Model-based design and analysis of life table experiments for insect vectors

Kamil Erguler[1*], Murat Can Demirok[2], Filiz Gunay[2], Mina Petrić[3,4,5], Mihaela Kavran[3], Dušan Veljko Petrić[3], Bulent Alten[2].

1 The Cyprus Institute, Climate and Atmosphere Research Excellence Centre (CARE-C), 20 Konstantinou Kavafi Street, 2121, Aglantzia, Nicosia, Cyprus
2 Hacettepe University, Faculty of Science, Biology Department, Ecology Division, VERG Laboratories, 06800, Beytepe-Ankara, Turkey
3 University of Novi Sad, Faculty of Agriculture, Laboratory for Medical and Veterinary Entomology, 21000 Novi Sad, Serbia
4 Avia-GIS NV, 2980 Zoersel, Belgium
5 Department of Physics, Faculty of Sciences, University of Novi Sad, 21000 Novi Sad, Serbia

*** k.erguler@cyi.ac.cy**

Abstract

Life tables help identify physiological differences in distinct development stages and detect potential vulnerabilities for conservation and control. However, cataloguing mortality, development, and fecundity by following each individual could be challenging in insects due to interweaving generations and development stages.

Here, we propose to use age- and stage- structured population dynamics modelling to help derive life table characteristics from the observed dynamics of reared populations. We examine a hypothetical case, a simulated population with known life parameters, and two experimental cases, observations of the population dynamics of the mosquito vector *Culex quinquefasciatus* and *Culex pipiens*, to demonstrate that model-based inference can correctly identify life parameters from longitudinal observations. The analysis reveals not only the differential physiological behaviour of distinct development stages, but also identifies the degree to which each parameter can be inferred from the data.

The methods introduced constitute a model-based approach to identifying life table characteristics from incomplete longitudinal data, and help to improve the design of life table experiments. The approach is readily applicable to the development of climate- and environment-driven population dynamics models for important vectors of disease.

Keywords: Insect, vector, population dynamics, age-structured, stage-structured, approximate, Bayesian, inference, life table, epidemiology.

Introduction

Life tables provide valuable insights into the dynamics and environmental dependence of populations [3, 28]. They have been used extensively for species conservation by identifying sensitive life stages [44] and key environmental drivers [51], and for designing effective control strategies against pests and disease vectors [3, 38, 40]. Life tables have become an essential component of mathematical models representing population dynamics by providing the biological foundations for model development [17, 41, 50].

A common practice of constructing life tables involves following a cohort of individuals in time and cataloguing a set of key life processes including mortality, development, and fecundity [3]. Typically, factors such as age-specific survivorship (the number of females or males on any day) are recorded and net reproductive rate (the expected number of females produced by the females of a cohort) is estimated. However, the list is adapted to fit the special needs of the species being examined [2].

Mortality is a key component of life tables; however, for many insect vectors, including mosquitoes, ticks, and small biting flies, it is difficult to assess the viability of certain stages until development commences to the next stage. For instance, egg viability is commonly quantified through larval hatching, *i.e.* eggs which do not develop into larvae after an appropriate amount of time are assumed dead [4, 19, 37].

Often, subsequent development stages overlap resulting in variability being amplified as development progresses. It might be impractical to isolate individuals or technically impossible to tag and follow each individual throughout development. Such interventions could impact development and survival, and result in inaccurate derivations [20, 47]. Multiple cohorts could be constructed to study the characteristics of different stages; however, this could restrict the experimental setup [12, 24].

Mathematical models have long been used to study and improve the understanding of many physical, chemical, and biological processes [5, 13]. By definition, a model can only represent a simplified version of the reality, and it is informed by prior knowledge, which comprises past observations, relevant literature, and expert opinion [36]. Canonical modelling approaches capture population heterogeneity partially by incorporating discernible development stages as distinct entities, *i.e.* stage-structured models [11, 27, 31]. Only a subset incorporates time-delay to represent time-driven heterogeneity as a result of, for instance, differential development rates [29, 39, 42]. By propagating age-structure in a difference equations framework, the sPop modelling framework offers an alternative to represent heterogeneity in natural populations [14].

The age-structured discrete time population dynamics model implemented in sPop is an extension of the works of Leslie [32] and Lefkovitch [11, 31]. As specified by the model, individuals are grouped into distinct stage classes but retain their developmental ages, *i.e.* the level of completion of their development. Thus, individuals in each age-development group are subject to the same survival, developmental, and reproductive characteristics. The model requires programmatic handling of a dynamic set of difference equations, and this is provided by the sPop package implemented in C, Python, and R [14]. The approach was first introduced by Erguler *et al.* (2016) [17] for modelling the environment-driven population dynamics of *Aedes albopictus*. Erguler *et al.* demonstrated in subsequent work that age- and stage-structured population models can be useful in representing field populations and disease transmission [15, 16]. This is achieved by inferring appropriate life table characteristics with regard to field observations and large-scale meteorological datasets in the absence of microenvironment conditions.

In this study, we elaborate on the model-based inference approach to identify life table characteristics under controlled laboratory settings. We follow common experimental practices where observations of viability and stage transformations are recorded at regular time intervals to the extent possible. We investigate the

development of a typical insect and aim to derive parameters pertaining to life characteristics at different stages. We propose that mathematical models and methods of inference can be used to derive these parameters and to relax, to a certain extent, the constraints on developing more informative but complex study designs.

Model-based inference using approximate Bayesian computation

The model-based inference framework we propose involves model construction, parameter optimisation, and Bayesian posterior sampling. More specifically, a combination of a model and a particular parameter configuration is offered as a potential representation of a life table experiment, and the value of this combination is assessed with respect to prior belief and the ability to replicate new observations.

In the Bayesian context, the prior probability of a model or a parameter configuration is informed by expert opinion and the existing knowledge base. The posterior probability, on the other hand, is informed by a novel set of observations in addition to the prior. According to this,

$$\Pr(\theta|\delta) \propto \Pr(\delta|\theta)\Pr(\theta), \tag{1}$$

where $\Pr(\theta)$ is the prior probability of the parameter configuration θ and $\Pr(\delta|\theta)$ is the likelihood, *i.e.* the probability of making the observation δ by simulating the model with θ [30, 48].

The prior probabilities can be the same among all proposals or set in favour of a particular hypothesis. The likelihood, however, is not often straightforward to define or evaluate. Deterministic models require explicit accounts of observational/intrinsic noise and time-correlation, which might be evident in the data. Although stochastic models can account for the noise, the estimation of the likelihood is, in most cases, computationally expensive. In response, approximate methods have been developed [35, 49]. In approximate Bayesian computation (ABC), likelihood is replaced with a simulation-based distance function, f, and the posterior is approximated with a given threshold ϵ. As a result,

$$\Pr(\theta|\delta) \approx \Pr(\theta|f(\delta, y(\theta)) \leq \epsilon), \tag{2}$$

which tends to $\Pr(\theta|\delta)$ as $\epsilon \to 0$. Here, f is the distance between the data and the corresponding model output $y(\theta)$.

Despite the use of approximation, exploring the entire posterior distribution is often computationally demanding, particularly when the number of parameters is large [34]. In order to reduce the computational demand, we propose generating posterior samples around the optimum posterior parameter. In our previous work, we referred to this partial posterior sample as the posterior mode and demonstrated that it could be useful for model development especially when competing models are complex [15–17].

A posterior mode, Θ, is defined as a model-parameter combination which includes a subset of all possible parameter configurations [17]. In the context of ABC, the approximate marginal probability of a posterior mode can be derived from Eqns. 1 and 2 as

$$\Pr(\Theta|f(\delta, y(\theta_\Theta)) \leq \epsilon) \propto \int \mathbb{1}(f(\delta, y(\theta_\Theta)) \leq \epsilon)\Pr(\theta_\Theta)d\theta_\Theta,$$

where θ_Θ represents a parameter configuration from Θ and $\mathbb{1}$ represents an indicator function returning 1 if $f(\delta, y(\theta_\Theta)) \leq \epsilon$ and 0 otherwise.

In order to identify the optimum value of the posterior, we propose using an optimisation algorithm on the likelihood, if available, or on the distance function. Due

to its simplicity and broad applicability, including deterministic and stochastic contexts, here, we employed the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm [7, 26]. In S1 Text, we present a sample R script to perform parameter optimisation with MCMC.

Once an optimum parameter configuration is identified, sampling can be performed using any posterior sampling algorithm, such as the canonical rejection method or MCMC [6]. Toni *et al.* (2008) developed a sequential Monte Carlo algorithm (ABC-SMC) to sample from the approximate posterior distribution [49]. ABC-SMC is powerful, widely applicable, and straightforward to implement.

Here, we adapted the ABC-SMC algorithm to yield samples predominantly from a posterior mode. We achieved this by initiating the algorithm at an optimum parameter configuration and executing it with a fixed distance threshold. Eventually, a stationary distribution emerges which represents a sample from the posterior mode around the optimum. It is possible to obtain such a sample from a single MCMC chain originating from the same optimum provided that the chain is sufficiently iterated. However, the ABC-SMC algorithm provides a computationally feasible alternative [49]. In S2 Text we present a sample script to run the algorithm in R.

Simulated life table experiment

In order to demonstrate the approach, we hypothesised a typical life table experiment which starts with a given number of mosquito eggs, and the eggs are observed to develop gradually into larva and pupa. In order to represent this experiment, we used a stochastic sPop model with fixed daily survival and gamma-distributed development times. Through this model, we assumed that development rate is time-dependent, *i.e.* the probability of completion increases gradually as time passes, and the individuals at the same developmental age can be grouped together.

We present in S3 Text an implementation of this model in R. The following is the master equation for eggs, larvae, and pupa, *i.e.* the mathematical representation of the dynamics of their joint probability distribution.

$$
\begin{aligned}
& \Pr(Egg_t = x, Larva_t = y, Pupa_t = z) \\
= \; & \Pr(Egg_{t-1} = x+1, Larva_{t-1} = y, Pupa = z)\,(x+1)p_E \\
& \Pr(Egg_{t-1} = x, Larva_{t-1} = y+1, Pupa = z)\,(y+1)p_L \\
& \Pr(Egg_{t-1} = x+1, Larva_{t-1} = y-1, Pupa = z)\,(x+1)(1-p_E)d_E \\
& \Pr(Egg_{t-1} = x, Larva_{t-1} = y+1, Pupa = z-1)\,(y+1)(1-p_L)d_L \\
- \; & \Pr(Egg_t = x, Larva_t = y, Pupa_t = z)\;[xp_E + yp_L + x(1-p_E)d_E + y(1-p_L)d_L] \\
\text{and} & \\
& \Pr(Egg_0 = x_0, Larva_0 = y_0, Pupa_0 = z_0) = 1,
\end{aligned}
$$

where x_0, y_0, and z_0 represent the initial conditions and p and d represent the daily probabilities of mortality and development, respectively. The model implies that eggs and larvae survive at fixed daily probabilities $(1-p)$, eggs and larva development occurs with a daily probability of d, given survival, and larvae accumulate as pupae when they complete development. While survival is independent of age, the daily development probability is age-dependent and follows a gamma distribution with a given mean (μ_d) and standard deviation (σ_d).

In Figure 1, we present three realisations of this model simulated with an initial condition $(x_0, y_0, z_0) = (100, 0, 0)$ and an arbitrary set of parameter values ($p_E = 0.01$, $\mu_{d_E} = 25$, $\sigma_{d_E} = 15$, $p_L = 0.05$, $\mu_{d_L} = 10$, and $\sigma_{d_L} = 5$). In the context of this hypothetical case, these model parameters can be regarded as life table parameters.

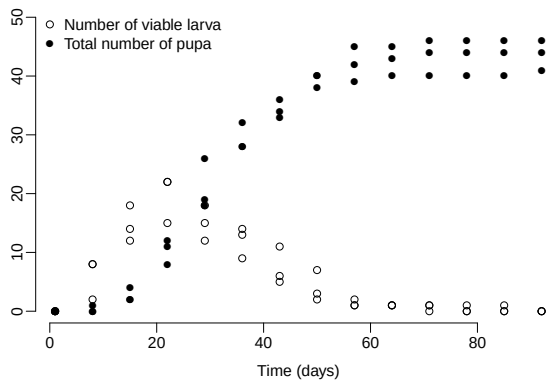


Figure 1. Simulated data of a typical life table experiment. The number of viable larvae (hollow circles) and the total number of pupa emergence (solid circles) are shown for each day. Three replicates were simulated starting with 100 eggs using the stochastic sPop model as described in the main text.

We proposed four models as potential explanations for the data in Figure 1. We investigated each one separately and examined their advantages and disadvantages in identifying life table parameters. We employed a generic score function, the sum of squared errors (SS), as the distance function.

$$SS = f(\delta, y(\theta)) = \sum_t (\delta_t - y_t(\theta))^2. \tag{3}$$

sPop: age- and stage-structured difference equations model

To begin with, we proposed a deterministic sPop model representing the egg and larva stages with fixed daily survival and gamma-distributed development times. The products of developed larvae are accumulated in the third stage as pupa similar to the stochastic version explained above.

The difference equations describing the model are given below, while a typical implementation of the model in R is given in S4 Text.

$$\begin{aligned} Egg_{t+1} &= (1 - p_E - d_E)\, Egg_t \\ Larva_{t+1} &= (1 - p_L - d_L)\, Larva_t + d_E\, Egg_t \\ Pupa_{t+1} &= Pupa_t + d_L\, Larva_t \end{aligned}$$

where p and d represent daily fraction of mortality and development, respectively. The model employs 6 parameters in total: p_E, μ_{d_E}, σ_{d_E}, p_L, μ_{d_L}, and σ_{d_L}, where μ and σ represent the mean and standard deviation, respectively, of the corresponding gamma-distributed development time.

Diff: stage-structured difference equations model

Lefkovitch in 1965 [31] introduced a generic stage-structured population dynamics model by grouping individuals into distinct stage classes with identical life table characteristics. In a modelling study for loggerhead turtles, Crouse *et al.* (1987) described a method for estimating the projection matrix, which quantifies daily change in each stage and between stages, using life tables [11].

The model, which is referred to as the Diff model in this context, permits a heterogenous stage class where the age of individuals are distributed as a function of survival probability (q) and development time (d); however, instead of iterating this distribution, which is essentially done in sPop, daily mortality and transition rates are estimated for a stage by aggregating the distribution.

In essence, the following rates are estimated:

- The daily fraction of individuals surviving and growing into the next stage:

$$G = \frac{(1-q)\, q^d}{1 - q^d}. \tag{4}$$

- The daily fraction of individuals surviving and remaining in the same stage:

$$P = \frac{q - q^d}{1 - q^d}. \tag{5}$$

The deterministic difference equations describing the Diff model are given below, and a typical implementation of the model in R is presented in S5 Text.

$$\begin{aligned} Egg_{t+1} &= P_E\, Egg_t \\ Larva_{t+1} &= P_L\, Larva_t + G_E\, Egg_t \\ Pupa_{t+1} &= Pupa_t + G_L\, Larva_t. \end{aligned}$$

The model represents eggs progressing into larvae, which subsequently develop into, and accumulate as, pupae. In total, the Diff model employs 4 parameters: p_E, d_E, p_L, and d_L, where mortality ($p = 1 - q$) is replaced with survival in concert with the sPop model.

ODE: ordinary differential equations model

The use of ordinary differential equations (ODEs) in vector population and disease transmission modelling dates back to the early twentieth century [43, 45]. More recently, ODE models have been developed to represent climate-driven population dynamics of many vector species [8, 50].

Despite their prevalence and relative ease of use, ODE models poorly represent population heterogeneity and the time-delay caused by underlying biological processes. Nevertheless, we followed the established practices and developed a deterministic population dynamics model to represent the transition of eggs to larvae, and then, to pupae.

The equations describing the ODE model are given below, and a typical implementation of the model in R using the deSolve package is presented in S6 Text.

$$\begin{aligned} \tfrac{d}{dt} Egg &= -(\alpha_E + \beta_E)\, Egg_t \\ \tfrac{d}{dt} Larva &= \alpha_E\, Egg_t - (\alpha_L + \beta_L)\, Larva_t \\ \tfrac{d}{dt} Pupa &= \alpha_L\, Larva_t, \end{aligned}$$

where α and β represent instantaneous development and mortality rates, respectively. In total, the model employs 4 parameters: α_E, β_E, α_L, and β_L.

DDE: delay differential equations model

Nisbet, Gurney, and Lawton developed a stage-structured delay differential equations (DDEs) model to represent insect population dynamics [25, 39]. The authors employed instantaneous rates of mortality and fecundity, and incorporated time-delay in the development process. Recently, Kettle *et al.* implemented the model in R, which is available in the stagePop package [29].

The model enables dynamic variation in development times; however, it assumes that all individuals in a stage behave in exactly the same way regarding survival and development. This implies a homogenous stage where a sharp transition to the next stage accrues from a cohort of individuals. As a potential workaround for representing developmental variation in a stage, the authors proposed inventing a plausible population history for some time before zero until the required (or observed) variation builds up [25, 39].

Here, we employed a parameter, τ_t, to control the rate of entry to the initial stage. In this context, variable entry times can be seen as variable starting times for development. In turn, one observes a seemingly gradual completion of development and

a smooth transition between subsequent stages. Although the authors recommended generating the population history some time before $t = 0$, in this context, variable entry would yield a similar behaviour but a shift in development times. For instance, the minimum and maximum development times for variable entry are $t_d - \tau_t$ and t_d, respectively, where t_d is the time it takes to the first completion of development. Conversely, the minimum and maximum development times for the population history approach are t_d and $t_d + \tau_t$, respectively.

The consequence of either of the approaches is that each individual follows exactly the same development process, but from a different time frame. At any instant, while some individuals are commencing development, others might have already finished it. Development always takes the same time. In cases of population interventions, such as blood feeding, or when different stages must have different developmental age distributions, this workaround may not be applicable. Nevertheless, in this context where only two stages are of interest, the approach seems applicable.

In order to represent a desired distribution of heterogeneity, *e.g.* gamma, negative binomial, or Gompertz [21], a customised rate of entry can be constructed from the derivative of the distribution function. A simple first-order approximation is to assume that individuals arrive at a constant rate, which results in a uniformly distributed transition time. Here, we adopted a constant rate of entry to examine its impact on the model and fit to the data.

A typical implementation of the DDE model in R using the stagePop package is given in S7 Text. The equations describing the model can be written as

$$\begin{array}{lcl} \frac{d}{dt}Egg & = & I_t - M(\alpha_E, Egg) - \beta_E\, Egg_t \\ \frac{d}{dt}Larva & = & M(\alpha_E, Egg) - M(\alpha_L, Larva) - \beta_L\, Larva_t \\ \frac{d}{dt}Pupa & = & M(\alpha_L, Larva), \end{array}$$

where α and β represent per capita development and death rates, respectively, and M is a function of α and the historical state of the corresponding stage. In this equation, I_t represents the rate of entry to the egg stage, and it is given by

$$I_t = \begin{cases} Egg_0/\tau_T & \text{if} \quad t < \tau_T \\ 0 & \text{else} \end{cases} .$$

where Egg_0 is the desired size of the initial egg population and τ_T is the time allocated to reach Egg_0. Consequently, τ_T controls the variability of development in the egg population. In total, the model has 5 parameters: α_E, β_E, α_L, β_L, and τ_T.

Results and discussion

As a result of the optimisation-posterior sampling procedure, described in Model-based inference using approximate Bayesian computation, we observed that the four models offered different explanations to the data. We list the distance thresholds (ϵ) and minimum distances achieved by each model in Table 1. In Figure 2, we present the output ranges of each model in comparison with the observations.

Table 1. Model performances against the simulated life table experiment.

Model	Threshold (ϵ)	Min. distance
sPop	362.471	331.4
Diff	1041.56	1031
ODE	1118.7	1048
DDE	558.181	535.8

As a result, we observed that the sPop model performed better than the other three in offering an explanation to the observations. The sPop and target models exhibit significant similarities; their main difference is the omission of stochasticity in the former. As a consequence, the observed performance of the model was expected.

As the second best explanation, the DDE model achieved distances comparable to those of sPop, and performed far better than the Diff and ODE models. The results suggest that the uniform distribution in entry times, employed for the DDE model, contributed to explaining the observed gradient in stage transition.

Finally, we observed that the Diff and ODE models failed to represent the time delay emerging from development, and performed worse than the others. Both models behaved similarly and yielded a consistent transition rate from the egg stage to the larva and pupa stages. This emphasises that the discrete projection matrix and continuous ODE approaches differ slightly, but mostly in propagating different amounts of numerical error during simulation. The use of transition rates, P and G in Eqns. 4 and 5, merely offers a way to approximate stage transition parameters from life table parameters.

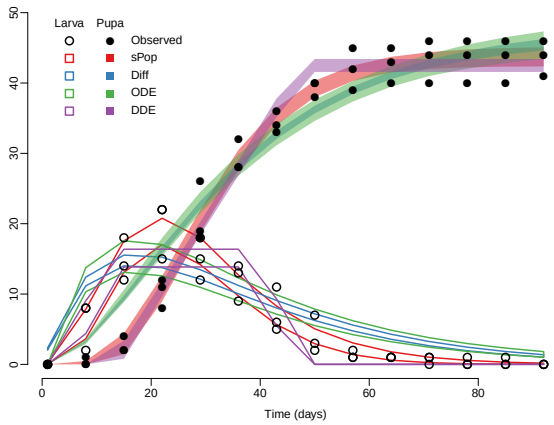


Figure 2. Comparison between the observations and model fits. Shaded and outlined trajectories display the ranges of pupae and larvae, respectively, as simulated using one of the four models and its corresponding 1000 posterior parameter samples. The observations are shown as points as in Figure 1.

Offering a deterministic model as an explanation to a stochastic system (or a deterministic system with observation error) leaves the noise in observations unaccounted for. This results in a non-zero distance value as the minimum; *i.e.* even the best models achieve a distance no lower than a certain value. Although raising the complexity of a model may result in better fitting, exceedingly complex models risk overfitting, which leads to noise being misinterpreted as signal. Here, no model-parameter combination achieved a distance lower than 331.4.

Learning from the data with a model. The inference process offers insights into the observed system through the distribution of each posterior sample. The accuracy of these insights depends heavily on model structure and the plausibility of its underlying assumptions. Nevertheless, a slight shift in parameter space should be expected due to the small sample size and the intrinsic stochasticity of the observed system. This leads to the theoretical optimum value of the posterior to deviate around the target value. The magnitude of this deviation decreases with increasing sample size.

In Table 2, we present the parameter configurations achieving minimum distances in each posterior sample. We observed that the inference resulted in close proximity of the target configuration with the sPop model. The Diff model, on the other hand, implied longer development times for both eggs and larvae. Egg mortality was higher, but larva mortality was lower than expected. The ODE model incorporates instantaneous transition rates, which could roughly translate into exponential decay rates. According to this, the model implied that the expected development time for eggs was $1/0.0271 \approx 37$ days and for larvae was $1/0.0648 \approx 15$ days. Although egg mortality was close to the expected value, the model implied lower mortality for larvae.

According to the DDE model, the egg mortality is close to the expected value, however, larva mortality is higher than the target. Interpretation of the development times is less straightforward. The model assumes identical development times for all individuals. When the variable entry modification is not applied, see DDE: delay differential equations model, and the model is provided with a development time of, for instance, 25 days, all individuals are incubated for 25 days in the egg stage and transferred to larvae at once. This clearly contradicts the gradual transition seen in the

observations. When we applied the variable entry modification, we observed that egg and larva development took approximately 5 and 8 days, respectively, and the observed variation was due to individual eggs commencing development around 35 days apart (Table 2).

Table 2. Best-fitting model-parameter configurations. Models and representative minimum distance parameter values are shown as possible explanations to the observations. The target model and parameter values are given in the final row.

Model	**Parameters (name - value)**											
sPop	p_E	0.0116	μ_{d_E}	25.5	σ_{d_E}	14.3	p_L	0.0547	μ_{d_L}	9.74	σ_{d_L}	5.19
Diff	p_E	0.0244	d_E	30.2			p_L	3.31×10^{-5}	d_L	15.1		
ODE	β_E	0.0121	α_E	0.0271			β_L	0.0291	α_L	0.0648		
DDE	β_E	0.0112	α_E	5.08			β_L	0.0995	α_L	8.04	τ_T	34.6
sPop (stoch.)	p_E	0.01	μ_{d_E}	25.0	σ_{d_E}	15.0	p_L	0.05	μ_{d_L}	10.0	σ_{d_L}	5.0

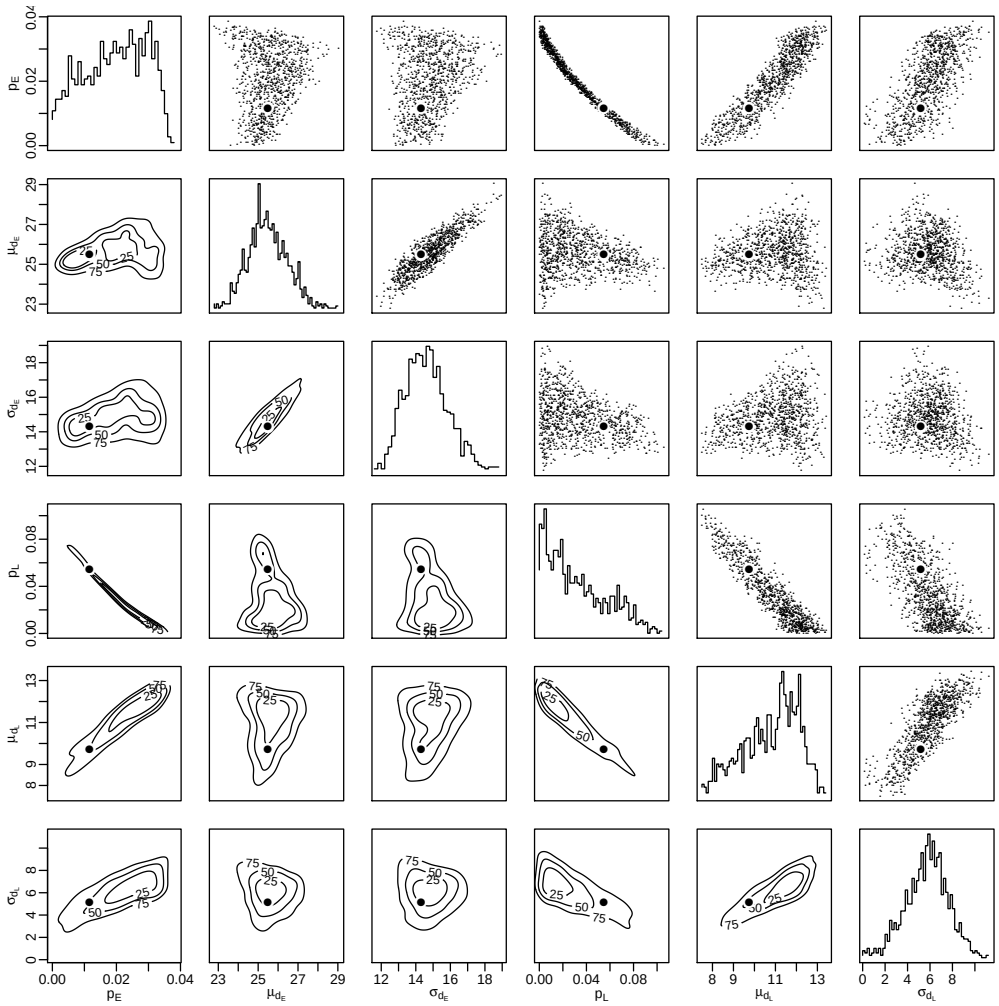


Figure 3. Posterior mode for the sPop model. The figure shows 1000 posterior samples for sPop obtained around the optimum point given in Table 2. Parameter values are shown above, kernel density estimates are plotted below, and marginal frequency distributions are given along the diagonal. The optimum parameter configuration is marked with a solid circle.

The developmental age distribution constructed with variable entry becomes identical for both the egg and larva stages. When this variation is integrated into development times, the model can be interpreted as representing a population of fast and slowly developing individuals. For instance, we observed that a small fraction of the population completed both stages in about 13 days, while a disparate fraction completed development slowly in approximately 48 days. In essence, larva development took about 1.6 times longer than egg development. Consequently, the maximum development times were approximately 18 days for eggs and 30 days for larvae.

Posterior samples obtained with ABC are weighed equally. When sampling with ABC, any parameter with a distance lower than the threshold is treated equally and assigned the same weight when calculating model predictions [49]. The optimum parameter configuration does not represent the single best fit to the data. The entire posterior should, therefore, be explored and all relevant parameter configurations should be included in predictions. This is often not practical or computationally feasible. Here, we focus on a partial posterior sample, the posterior mode, which is sampled around an optimum value [17].

We plot the posterior samples obtained for the sPop model in Figure 3 together with the marginal frequency distributions and kernel density estimates (KDEs, obtained with the `kde` function of the `ks` package in R). The posterior samples for the remaining three models are given in Figures S2 Figure for the Diff model, S3 Figure for the ODE model, and S4 Figure for the DDE model.

A strict negative correlation is evident between the egg and larva mortalities in the posterior samples of the sPop model (Fig. 3). This indicates that the combination of low egg mortality and high larva mortality has the same effect as the opposite configuration. In the vicinity of the target values for mortality, we observed that the samples included all target values for survival and development.

A negative association between the mortalities of egg and larva was also evident in the posterior samples of the DDE model (see S4 Figure). The target values corresponding to the parameterisation of DDE are $\beta_E = 0.01$ and $\beta_L = 0.05$. We observed that with $\epsilon = 558.181$, the samples excluded these values. Similarly, the samples obtained for the ODE and Diff models (with $\epsilon = 1118.7$ and $\epsilon = 1041.56$, respectively) did not extend over the corresponding target values ($p_E = \beta_E = 0.01$ and $p_L = \beta_L = 0.05$) with the exception of β_E ranging between 0.008 and 0.014 for the ODE model.

The corresponding target values for development times are $d_E = 25$ and $d_L = 10$ for the Diff model and $\alpha_E = 0.04$ and $\alpha_L = 0.1$ for the ODE model. We observed that the posterior samples of either model included these values: samples for Diff ranged $d_E = 28.28 - 32.30$ and $d_L = 14.34 - 15.88$, and for ODE ranged $\alpha_E = 0.024 - 0.032$ and $\alpha_L = 0.057 - 0.076$. We note that the inferred development times were consistently longer than expected when using either the Diff or the ODE models.

Samples for the DDE model implied that egg development time ranges between 2.30 and 6.29 days, while larva development time ranges between 7.32 and 11.26 days. This is accompanied by the variable entry time ranging between 32.94 and 36.23 days. We incorporated variable entry times into development times and focused on total development to facilitate model comparison. According to this, the DDE model implied a median development time in the range of 29.84-31.39 days, however, the target value was 35 days.

New evidence leads to better inference. It is possible to further constrain the posterior by adding new observations. For instance, the difference between the total larva and pupa production can be used to pinpoint mortality rates for eggs and larvae.

In order to incorporate new data, we re-configured Eqn. 1 by considering the posterior $\Pr(\theta|\delta)$ as a prior with regards to new data, δ'. Thus,

$$\Pr(\theta|\delta') \propto \Pr(\delta'|\theta)\Pr(\theta),$$

where

$$\Pr(\theta) = \Pr(\theta|\delta) \approx \Pr(\theta|f(\delta, y(\theta)) \leq \epsilon).$$

We replaced the likelihood, $\Pr(\delta'|\theta)$, with a distance function, f', and sampled with an arbitrary threshold, ϵ', as in Eqn. 2.

$$\Pr(\theta|\delta') \approx \Pr(\theta|f'(\delta', y'(\theta)) \leq \epsilon').$$

We assigned the new distance function as

$$f'(\delta', y'(\theta)) = |\delta' - y'(\theta)|,$$

where $y'(\cdot)$ is the total number of dead larva predicted by the proposed deterministic model with θ and $\epsilon' = 6$.

We observed that for the simulation shown in Figure 1, δ' is, on average, 31 (the values were 25, 31, and 37).

We used the 1000 samples obtained for the sPop model as samples from the prior distribution. We observed that 150 of them yielded dead larvae in the range 25-37, which conforms the distance threshold $\epsilon' = 6$. The samples, as shown in S1 Figure, constrained the posterior distribution for both the egg and larva mortalities around the corresponding target values. The additional datum also constrained the inferred development times to 24.07-26.92 ($\pm$12.48-16.43) for eggs and 8.16-11.67 ($\pm$0.33-8.47) for larvae.

Inference with experimental datasets

Life table parameters for *Culex quinquefasciatus*

Culex quinquefasciatus, commonly known as southern house mosquito, is an important vector of bancroftian filariasis and West Nile virus, and it has been shown capable of transmitting chikungunya and Zika viruses [1, 22, 46]. It is widely distributed across the tropics and sub-tropics.

In 2010, Gunay *et al.* reported a comprehensive study of *Cx. quinquefasciatus* (Say, 1823) [24] at five constant temperatures (15, 20, 23, 27, and 30°C). As described in Gunay *et al.* 2009 and 2010 [23, 24], the study involved regular observations of a total of 750 1st instar larvae incubated at 65$\pm$5% relative humidity and 14:10 hours of light:dark (L:D) cycle for three consecutive generations in three replicates.

Although the experiment lead to the derivation of several life table parameters, such as adult longevity, generation times, and egg-to-adult survival rates, characteristics of intermediate development stages could not be accurately determined due to overlapping stage durations. Here, we used a subset of these observations (second generation larvae developing at 15°C) to demonstrate the added value of model-based analysis in resolving development and survival characteristics of larvae and pupae (Fig. 4a).

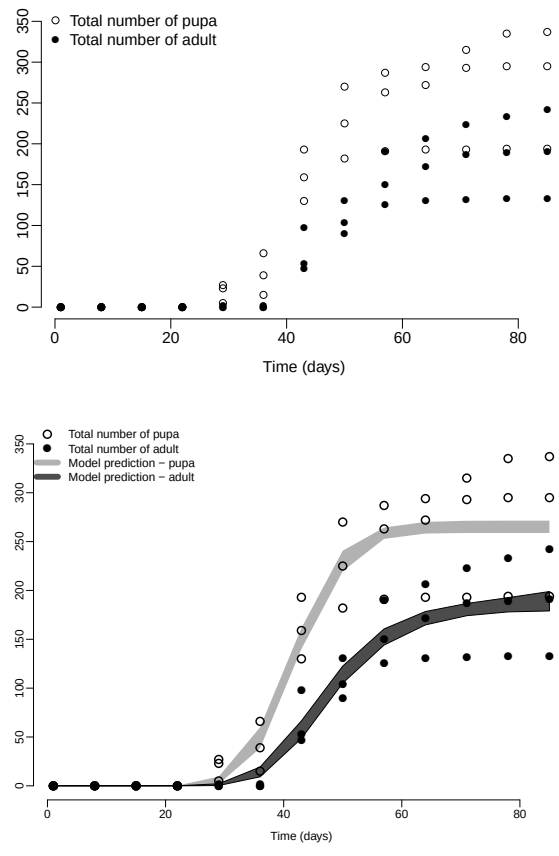


Figure 4. Life table experiment for laboratory-reared *Culex quinquefasciatus*. In (a), the total number of pupa (hollow circles) and adult mosquitoes (solid circles) emerging from 750 larvae are shown in 3 replicates. In (b), the observations are shown with model predictions obtained using the sPop model with 1000 posterior samples. The longitudinal dataset was adopted from Gunay *et al.* (2009) [23].

We developed a deterministic sPop model with two consecutive stages, larvae and pupae, and recorded the number of emerging adults similar to the model in section sPop: age- and stage-structured difference equations model. As a result of optimisation and sampling from the optimum posterior mode, we obtained 1000 parameter configurations, which yielded the predictions shown in Figure 4b. The model output seen in the figure corresponds to the expected trajectories of pupae and adults with regard to the three replicates obtained from the laboratory experiments.

We present the posterior samples in Figure 5. A strict positive correlation is evident in the figure between the mean and standard deviation of pupa development time ($\sigma = 2.146\mu - 9.028$). With the amount of data utilised, it was not possible to pinpoint pupa development, but, rather, to argue that its standard deviation is roughly twice as much as its mean.

In addition to the above dataset, Gunay *et al.* reported that on average 255 pupae and 190 adults were produced from 750 larvae at 15°C [23]. By using these observations and following the line of work described in the previous section, we found that 35 of the 1000 posterior samples yielded larva and pupa mortalities closer than 1% to the ones reported (Fig. 5). As a result, we observed that the daily mortality was constrained to 2.42±0.02% for larvae and 2.85±0.21% for pupae.

Gamma distribution with a standard deviation higher than its mean yields a low median. We observed that, although the mean development time of pupa was approximately 14.5 days, the median development time calculated using the inferred parameter values was 5.85±0.68 days. The probability of completing development at 15°C with respect to the time spent for the process are shown in Figure 6 for larva and pupa.

Figure 5. Posterior mode for the sPop model. The figure shows 1000 posterior samples for sPop obtained around the optimum point given in Table 2. The samples yielding larva and pupa mortalities closer than 1% to the observed are marked in red. Parameter values are shown above, kernel density estimates are plotted below, and marginal frequency distributions are given along the diagonal. The optimum parameter configuration is marked with a solid circle.

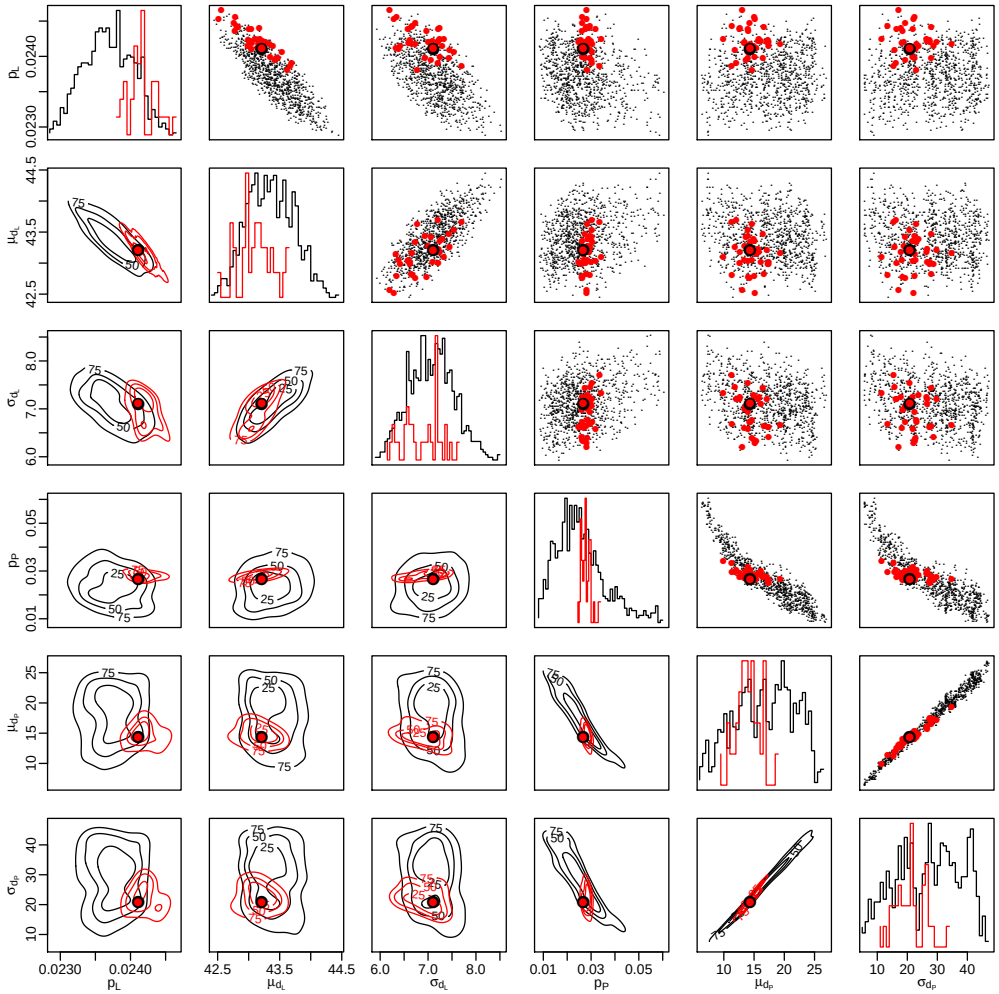
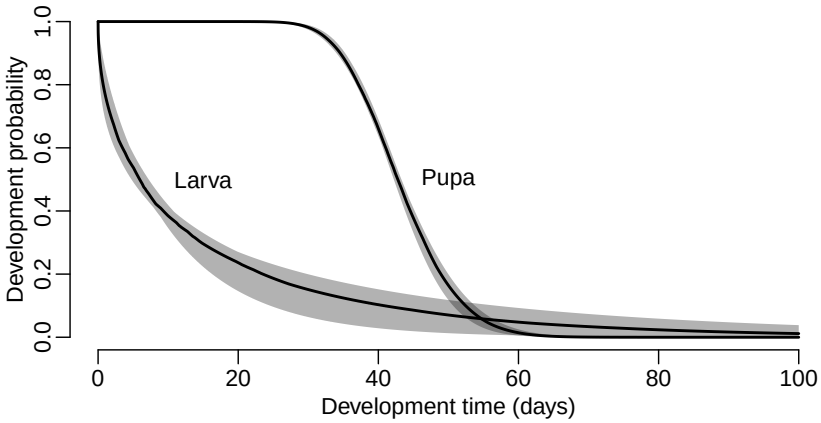


Figure 6. Larva and pupa development at 15°C for *Culex quinquefasciatus* inferred with the sPop model. Development probabilities are estimated using the sPop model and the 35 posterior samples as described in the main text. The medians are given as solid lines and the ranges are given as shades for larvae and pupa.



Life table parameters for *Culex pipiens*

In contrast to its southern neighbour, *Culex pipiens* is known as the northern house mosquito as it is widely distributed across the temperate countries in North America, Europe, Asia, and North and East Africa [9, 18]. It is an important vector of West Nile and St. Louis encephalitis viruses and filarial worms [18].

In 2019, a life table experiment was conducted by Petrić's group with a laboratory-reared population of *Culex pipiens* biotype molestus to examine its environmental dependence. Four batches of eggs (180, 177, 161, and 153 eggs) were prepared and their development was observed at room temperature. Each batch was kept in 1.5 L water at 19.9±1.0°C and 38.8±2.5% relative humidity. Following hatching and larva development, pupa production was recorded each day until all larvae either died or developed. We present the observations in Figure 7(a).

While pupa production was recorded, counts of viable eggs or larvae were not. Instead, the group noted that none of the eggs hatched on the second day and all the eggs hatched after four days. We marked the times of these two additional observations with arrows in the figure.

In order to represent this experiment, we developed a deterministic two-stage sPop model with egg and larva stages where pupa production is recorded daily. By default, sPop is a daily time-step difference equations model. However, due to the rapid hatching of the eggs, we employed a time unit of quarter days where each iteration accounts for 6 hours rather than 1 day. Following each simulation, we calculated the daily mean number of viable eggs and larva and the total pupa production for each day.

As the distance function, we employed the sum of squares distance for pupa production and included the two additional observations, *i.e.*

$$SS = \sum_t (\delta_t - P_t(\theta))^2 + L_{t1}(\theta)^2 + E_{t2}(\theta)^2,$$

where δ_t refers to the number of pupa produced on day t, $P_t(\theta)$ is the corresponding model simulation with parameter θ, $L_{t1}(\theta)$ is the simulated number of larvae on $t1$ (1 day after oviposition), and $E_{t2}(\theta)$ is the simulated number of eggs on $t2$ (4 days after oviposition).

Following the same optimisation and posterior sampling procedure as in previous sections, we observed that the model closely represented the bulk of observations on pupa production (Fig. 7(b)). Due to the lack of detailed records on earlier stages, a wide range of uncertainty was assigned for egg and larva development.

As an added value, the model offers a mean to drastically improve precision. A key observation is revealed upon close examination of the posterior mode. As seen in Figure 7(c), the first three days beginning with the observation of first egg hatching seem to offer the ability to resolve larva development time. If the mean development time (μ_{d_E}) is around 11 days, the peak in the number of larva realises later than if μ_{d_E} was around 13 days.

For instance, let us assume that on December 7th, the first day of egg hatching, the initial number of larvae were counted and found to be between 15 and 35 in each of the 4 replicates. This observation constrains both egg and larvae mean development times around $\mu_{d_E} = 2.64$ and $\mu_{d_L} = 12.2$ days, respectively while not significantly affecting the parameters for survival and standard deviation (p_E, p_L, σ_{d_E}, and σ_{d_L}).

Figure 7. Life table experiment for laboratory-reared *Culex pipiens* at room temperature. In (a), the total number of pupa emerging from the four batches of egg are shown. In (b), the observations are shown with model predictions (minimum and maximum pupa production per day) obtained using the sPop model with 1000 posterior samples. In (c), model outputs for larvae are plotted and marked for a range of μ_{d_E} values from the posterior.

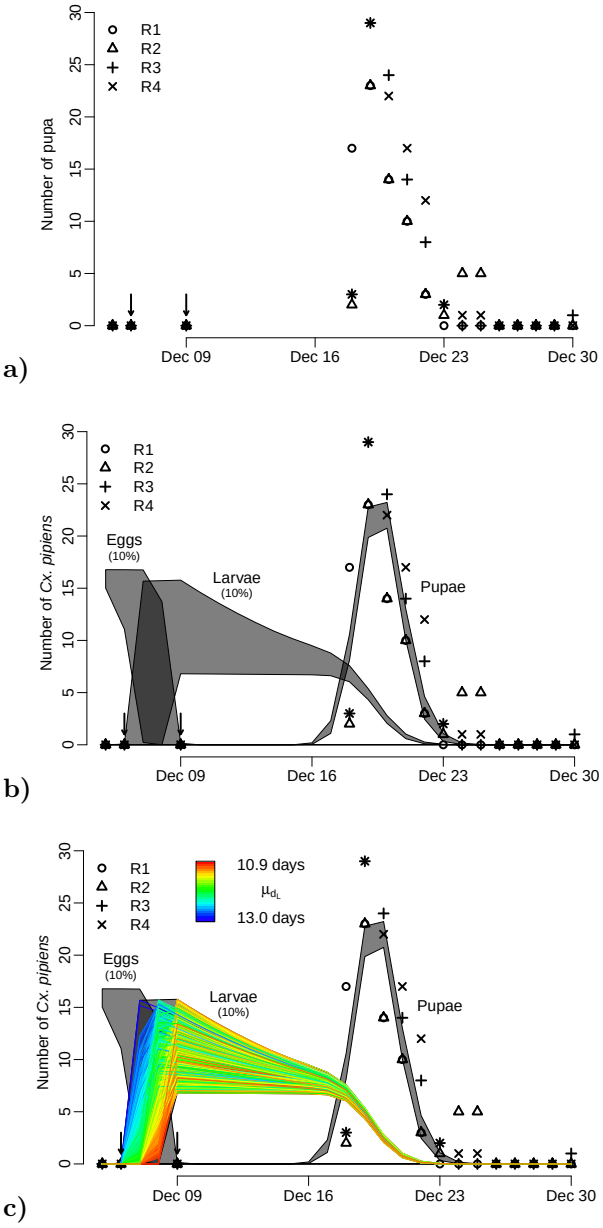
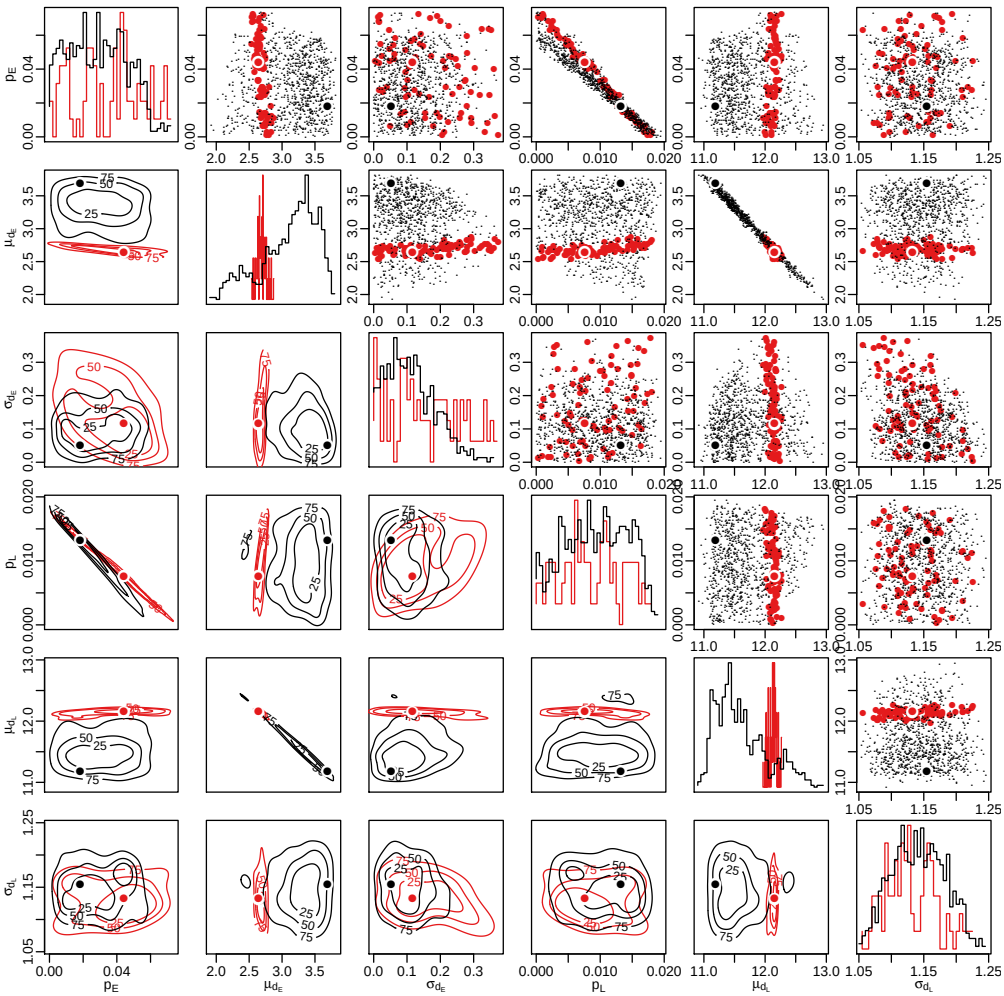


Figure 8. Posterior mode for the sPop model of *Culex pipiens*. The figure shows 1000 posterior samples, 88 of which are marked in red and correspond to a larva count of 15-35 on December 7th. Parameter values are shown above, kernel density estimates are plotted below, and marginal frequency distributions are given along the diagonal. The optimum parameter configuration is marked with two solid circles for each group.



Conclusion

Life tables have long been used to understand population dynamics and environmental dependence [10]. However, their resolution is limited given overlapping development stages, population heterogeneity, and the intricate interactions between individuals and their environment. Laboratory experiments under controlled conditions often risk disrupting such interactions, influencing natural behaviour, and affecting the outcome. We demonstrated that model-based inference can be a valuable tool in estimating life characteristics from complex and, often, incomplete observations while minimising the requirement of disruptive interventions.

We evaluated a representative set of frequently encountered tools readily accessible to the mathematical modellers of life sciences — more specifically, ecologists and epidemiologists [5]. We demonstrated that none of the selected mainstream modelling tools, except the sPop model, was appropriate to represent populations in required detail. We concluded that an appropriate model structure is necessary to derive valid conclusions. Model structure should be informed by valid assumptions arising from prior knowledge, and used to complement longitudinal observations.

In turn, appropriate models can provide valuable insights into the underlying biological processes and offer improvements in study design. Using the sPop model, we demonstrated that observations during stage transitions offer valuable discerning power on immature stage development characteristics. Focusing observations on such key time points could, therefore, provide sufficient information especially in limited resource cases.

We further note that, the variability in temporally aggregated observations (such as cumulative number of individuals) reaches a maximum towards the end. This introduces bias to the inference process especially when the variability is dismissed with the use of a generic distance function (such as the sum of squares). We suggest that further optimisations of this procedure should consider employing observations at singular time points, *i.e.* the number of individuals at time t, reducing the interval of observations, and using an appropriate distribution model, such as Poisson, as the distance measure.

Parameter values inferred from life table experiments are readily applicable as prior knowledge in population dynamics models. The resulting models can help to understand underlying biological processes including environmental dependence and disease transmission, and be used in projecting large-scale trends in climate-driven habitat suitability and disease risk.

Acknowledgments

This publication has been produced within the framework of the EMME-CARE project which has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 856612 and the Cyprus Government. The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein.

The work of KE, MCD, MP, MK, FG, DVP, and BSA was done within the framework of AIM-COST Action CA17108.

KE acknowledges the Cy-Tera Project (NEW INFRASTRUCTURE/STRATEGIC/0308/31), which is co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation, for providing the computational resources used in this research.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. N. Becker, D. Petric, M. Zgomba, C. Boase, M. Madon, C. Dahl, and A. Kaiser. *Mosquitoes and Their Control*, volume 57. 2010.

2. A. Belen and B. Alten. Variation in life table characteristics among Populations of Phlebotomus papatasi at different altitudes. *Journal of Vector Ecology*, 31(1):35–44, jun 2006.

3. T. Bellows and R. Van Driesche. Life Table Construction and Analysis for Evaluating Biological Control Agents. In *Handbook of Biological Control*, pages 199–223. Academic Press, 1999.

4. M. Q. Benedict. *Methods in Anopheles Research*. MR4 - BEI Resources, 2014.

5. F. Brauer and C. Castillo-Chavez. *Mathematical models in population biology and epidemiology*. Springer, 2 edition, 2010.

6. S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng. *Handbook of Markov Chain Monte Carlo*. 2011.

7. B. Calderhead. A general construction for parallelizing Metropolis-Hastings algorithms. *Proceedings of the National Academy of Sciences of the United States of America*, 111(49):17408–17413, dec 2014.

8. C. Christiansen-Jucht, K. Erguler, C. Y. Shek, M.-G. Basáñez, and P. E. Parham. Modelling anopheles gambiae s.s. population dynamics with temperature- and age-dependent survival. *International Journal of Environmental Research and Public Health*, 12(6):5975–6005, 2015.

9. A. T. Ciota and L. D. Kramer. Vector-virus interactions and transmission dynamics of West Nile virus. *Viruses*, 5(12):3021–3047, 2013.

10. D. Cox. Regression models and life tables. *J. R. Stat. Soc. B*, 34(2):187–220, 1972.

11. D. T. Crouse, L. B. Crowder, and H. Caswell. A stage-based population model for loggerhead sea turtles and implications for conservation. *Ecology*, 68(5):1412–1423, 1987.

12. H. Delatte, G. Gimonneau, A. Triboire, and D. Fontenille. Influence of temperature on immature development, survival, longevity, fecundity, and gonotrophic cycles of Aedes albopictus, vector of chikungunya and dengue in the Indian Ocean. *Journal of Medical Entomology*, 46(1):33–41, jan 2009.

13. Edda Klipp, W. Liebermeister, C. Wierling, and A. Kowald. *Systems Biology*. Wiley-VCH Verlag GmbH & Co. KGaA, 2016.

14. K. Erguler. sPop: Age-structured discrete-time population dynamics model in C, Python, and R. *F1000Research*, 7:1220, 2018.

15. K. Erguler, N. N. L. Chandra, Y. Proestos, J. Lelieveld, G. G. K. Christophides, and P. P. E. Parham. A large-scale stochastic spatiotemporal model for Aedes albopictus-borne chikungunya epidemiology. *Plos One*, 12(3):e0174293, 2017.

16. K. Erguler, I. Pontiki, G. Zittis, Y. Proestos, V. Christodoulou, N. Tsirigotakis, M. Antoniou, O. E. Kasap, B. Alten, and J. Lelieveld. A climate-driven and field data-assimilated population dynamics model of sand flies. *Scientific Reports*, 9(1):2469, dec 2019.

17. K. Erguler, S. E. Smith-Unna, J. Waldock, Y. Proestos, G. K. Christophides, J. Lelieveld, and P. E. Parham. Large-scale modelling of the environmentally-driven population dynamics of temperate aedes albopictus (Skuse). *PLoS ONE*, 11(2), 2016.

18. A. Farajollahi, D. M. Fonseca, L. D. Kramer, and A. Marm Kilpatrick. "Bird biting" mosquitoes and human disease: A review of the role of Culex pipiens complex mosquitoes in epidemiology. *Infection, Genetics and Evolution*, 11(7):1577–1585, 2011.

19. L. C. Farnesi, H. C. Vargas, D. Valle, and G. L. Rezende. Darker eggs of mosquitoes resist more to dry conditions: Melanin enhances serosal cuticle contribution in egg resistance to desiccation in Aedes, Anopheles and Culex vectors. *PLoS Neglected Tropical Diseases*, 11(10):1–20, 2017.

20. L. Gavotte, D. R. Mercer, R. Vandyke, J. W. Mains, and S. L. Dobson. Wolbachia infection and resource competition effects on immature Aedes albopictus (Diptera: Culicidae). *Journal of Medical Entomology*, 46(3):451–459, may 2009.

21. B. Gompertz. On the nature of the function expressive of the law of human mortality, and on the new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115(1825):513–583, 1825.

22. D. R. Guedes, M. H. Paiva, M. M. Donato, P. P. Barbosa, L. Krokovsky, S. W. Rocha, K. L. Saraiva, M. M. Crespo, T. M. Rezende, G. L. Wallau, R. M. Barbosa, C. M. Oliveira, M. A. Melo-Santos, L. Pena, M. T. Cordeiro, R. F. O. Franca, A. L. De Oliveira, C. A. Peixoto, W. S. Leal, and C. F. Ayres. Zika virus replication in the mosquito Culex quinquefasciatus in Brazil. *Emerging Microbes and Infections*, 6(8), 2017.

23. F. Gunay. *Farklı sabit sıcaklıklarda Culex quinquefasciatus (Diptera: Culicidae)'un reaksiyon normu ve kalıtsallığı.* Msc thesis, Hacettepe University, 2009.

24. F. Gunay, B. Alten, and E. D. Ozsoy. Estimating reaction norms for predictive population parameters, age specific mortality, and mean longevity in temperature-dependent cohorts of Culex quinquefasciatus Say (Diptera: Culicidae). *Journal of Vector Ecology*, 35(2):354–362, 2010.

25. W. S. C. Gurney, R. M. Nisbet, and J. H. Lawton. The Systematic Formulation of Tractable Single-Species Population Models Incorporating Age Structure. *Journal of Animal Ecology*, 52(2):479–495, 1983.

26. W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, apr 1970.

27. M. Iannelli and F. Milner. *The Basic Approach to Age-structured Population Dynamics Models, Methods and Numerics.* 2017.

28. A. M. Kakde, K. Patel, and S. Tayade. Role of Life Table in Insect Pest Management–A Review. *IOSR Journal of Agriculture and Veterinary Science*, 7(1):40–43, 2014.

29. H. Kettle and D. Nutter. StagePop: Modelling stage-structured populations in R. *Methods in Ecology and Evolution*, 6(12):1484–1490, 2015.

30. P. Kirk, T. Thorne, and M. Stumpf. Model selection in systems and synthetic biology. *Current Opinion in Biotechnology*, 2013.

31. L. P. Lefkovitch. The Study of Population Growth in Organisms Grouped by Stages. *Biometrics*, 21(1):1–18, 1965.

32. P. H. Leslie. On the Use of Matrices in Certain Population Mathematics. *Biometrika*, 33(3):183–212, 1945.

33. J. Liepe, C. Barnes, E. Cule, K. Erguler, P. Kirk, T. Toni, and M. P. H. Stumpf. ABC-SysBio-approximate bayesian computation in python with GPU support. *Bioinformatics*, 26(14):1797–1799, 2010.

34. J. Liepe, P. Kirk, S. Filippi, T. Toni, C. P. Barnes, and M. P. H. Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9(2):439–456, feb 2014.

35. P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324–15328, 2003.

36. R. May. Uses and Abuses of Mathematics in Biology. *Science*, 303(5659):790, feb 2004.

37. R. Menard. *Whole-genome analysis of plasmodium spp. Utilizing a new agilent technologies DNA microarray platform*, volume 923. Springer, 2013.

38. S. E. Naranjo and P. C. Ellsworth. The contribution of conservation biological control to integrated control of Bemisia tabaci in cotton. *Biological Control*, 51(3):458–470, 2009.

39. R. M. Nisbet and W. S. Gurney. The systematic formulation of population models for insects with dynamically varying instar duration. *Theoretical Population Biology*, 23(1):114–135, 1983.

40. G. R. A. Okogun. Life-table analysis of Anopheles malaria vectors: Generational mortality as tool in mosquito vector abundance and control studies. *Journal of Vector Borne Diseases*, 42(2):45–53, 2005.

41. J. A. Patz, W. J. Martens, D. A. Focks, and T. H. Jetten. Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environmental Health Perspectives*, 106(3):147–153, 1998.

42. G. Rosen. Time delays produced by essential nonlinearities in population growth models. *Bulletin of Mathematical Biology*, 49(2):253–255, 1987.

43. R. Ross. *Report on the prevention of malaria in Mauritius*. 1908.

44. H. Selina S. Application of Life-History Theory and Population Model Analysis to Turtle Conservation Author (s): Selina S . Heppell Published by : American Society of Ichthyologists and Herpetologists (ASIH) Stable URL : https://www.jstor.org/stable/1447430 REFERE. *Copeia*, 1998(2):367–375, 1998.

45. D. L. Smith, K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie. Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathog*, 8(4):e1002588, jan 2012.

46. R. Subra. Biology and control of Culex pipiens quinquefasciatus Say, 1823 (Diptera, Culicidae) with special reference to Africa. *International Journal of Tropical Insect Science*, 1(04):319–338, 1981.

47. H.-J. Teng and C. Apperson. Development and survival of immature Aedes albopictus and Aedes triseriatus (Diptera: Culicidae) in the laboratory: effects of density, food, and competition on response to temperature. *Journal of Medical Entomology*, 37(1):40–52, 2000.

48. T. Toni and M. P. H. Stumpf. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1):104–110, jan 2010.

49. T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of Royal Society Interface*, 6(31):187–202, dec 2008.

50. A. Tran, G. L'ambert, G. Lacour, R. Beno\ˆ\it, M. Demarchi, M. Cros, P. Cailly, M. Aubry-Kientz, T. Balenghien, and P. Ezanno. A rainfall- and temperature-driven abundance model for Aedes albopictus populations. *IJERPH*, 10(5):1698–1719, may 2013.

51. L. Xiushan, Z. Yalin, L. Youqing, and S. Josef. Life history , life table , habitat , and conservation of Byasa impediens (Lepidoptera : Papilionidae). *Acta Ecologica Sinica*, 26(10), 2006.

Supporting Information

S1 Text

The Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm (R script). The following code was inspired by the work of Darren Wilkinson (`https://darrenjw.wordpress.com/2010/08/15/metropolis-hastings-mcmc-algorithms/`).

```
mcmc <- function(pr,fun,lower,upper,kernel.sd,
                 niter=1000,thin=10,sig=1.0,verbose=FALSE) {
  acc <- 0
  vec <- NULL
  prn <- pr
  scrn <- scr <- fun(pr)
  for (n in 1:niter) {
      repeat {
          prn <- rnorm(length(pr),mean=pr,sd=kernel.sd)
          if (all(prn>=lower) && all(prn<=upper)) break
      }
      scrn <- fun(prn)
      if (!is.nan(scrn) && (scrn < scr || log(runif(1)) < (scr-scrn)/sig)) {
          pr <- prn
          scr <- scrn
          acc <- acc + 1
      }
      vecn <- c(scr,pr)
      if (n %% thin == 0) {
          vec <- rbind(vec,vecn)
          if (verbose) cat(sprintf("%d,%g,%s\n",n,acc,paste(vecn,collapse=",")))
          acc <- 0
      }
  }
  colnames(vec)<-c("score",sprintf("X%d",1:length(pr)))
  return(data.frame(vec,row.names=NULL))
}
```

S2 Text

The approximate Bayesian computation - sequential Monte Carlo (ABC-SMC) algorithm (R script). The following code was inspired by the works of Toni *et al.* (2008) and Liepe *et al.* (2010) [33, 49].

```
abc.smc <- function(pr,score,eps,size,lower,upper,kernel.sd,niter,verbose=TRUE) {
    scr <- score(pr)
    mat <- matrix(rep(c(1/size,0,scr,pr),size),nrow=size,byrow=TRUE)
    mat.new <- matrix(NA,nrow=nrow(mat),ncol=ncol(mat))
    for (i in 1:niter) {
        for (n in 1:nrow(mat.new)) {
            fcount <- 0
            repeat {
                v <- mat[sample(1:nrow(mat),1,prob=mat[,1],replace=TRUE),]
                pr.new <- rnorm(length(v)-3,mean=v[-(1:3)],sd=kernel.sd)
                scr <- score(pr.new)
                if (!is.nan(scr) && (scr<eps) &&
                    all(pr.new>=lower) && all(pr.new<=upper)) break
                fcount <- fcount + 1
            }
            if (verbose) print(c(i,n,fcount))
            w <- 1 / sum(apply(mat,1,function(v)
                 v[1]*exp(sum(dnorm(pr.new,mean=v[-(1:3)],sd=kernel.sd,log=TRUE)))))
            mat.new[n,] <- c(w,fcount,scr,pr.new)
        }
        mat.new[,1] <- mat.new[,1] / sum(mat.new[,1])
        mat <- mat.new
        if (verbose) {
            cat(paste(apply(mat,1,paste,collapse=","),collapse="\n"),sep="\n\n")
        }
    }
    return(mat)
}
```

S3 Text

The stochastic sPop model (R script). The model was used to simulate hypothetical experimental data of a given repetitions. The output of the function is a data.frame of weekly model output starting from day 1, the day after initiation.

```
fun.sPop.st <- function(pr,rpt=3) {
    require(albopictus)
    ret <- NULL
    for (r in 1:rpt) {
        print(r)
        egg <- spop(stochastic=TRUE,prob='gamma')
        larva <- spop(stochastic=TRUE,prob='gamma')
        add(egg) <- data.frame(number=100)
        larvad <- 0
        pupa <- 0
        for (n in 1:98) {
            iterate(egg) <- data.frame(death=pr[1],dev_mean=pr[2],dev_sd=pr[3])
            iterate(larva) <- data.frame(death=pr[4],dev_mean=pr[5],dev_sd=pr[6])
            larvad <- larvad + dead(larva)
            pupa <- pupa + developed(larva)
            add(larva) <- data.frame(number=developed(egg))
            if (n%%7==1) {
                ret <- rbind(ret,c(r,n,size(larva),larvad,pupa))
            }
        }
    }
    colnames(ret) <- c("rep","day","larva","larvad","pupa")
    ret<-data.frame(ret)
    return(ret)
}
```

S4 Text

The deterministic sPop model (R script). The output of the function is a data.frame of weekly model output starting from day 1, the day after initiation. The output comprise the following in the given order:

day	Day of observation (weekly)
stage2	Viable number of individuals in stage 2 at the time of observation
cum.stage2	Total number of individuals in stage 2 at the time of observation
cum.stage2.dead	Total number of non-viable individuals in stage 2 at the time of observation
cum.stage3	Total number of individuals in stage 3 at the time of observation

```
fun.sPop <- function(pr,tm,init) {
    require(albopictus)
    stage1 <- spop(stochastic=FALSE,prob='gamma')
    stage2 <- spop(stochastic=FALSE,prob='gamma')
    cum.stage2 <- 0
    cum.stage2.dead <- 0
    cum.stage3 <- 0
    add(stage1) <- data.frame(number=init)
    ret <- NULL
    for (n in 1:max(tm)) {
        iterate(stage1) <- data.frame(death=pr[1],dev_mean=pr[2],dev_sd=pr[3])
        iterate(stage2) <- data.frame(death=pr[4],dev_mean=pr[5],dev_sd=pr[6])
        new.stage2 <- developed(stage1)
        new.stage3 <- developed(stage2)
        dead.stage2 <- dead(stage2)
        add(stage2) <- data.frame(number=new.stage2)
        cum.stage2 <- cum.stage2 + new.stage2
        cum.stage2.dead <- cum.stage2.dead + dead.stage2
        cum.stage3 <- cum.stage3 + new.stage3
        ret <- rbind(ret,c(n,size(stage2),cum.stage2,cum.stage2.dead,cum.stage3))
    }
    colnames(ret) <- c("day","stage2","cum.stage2","cum.stage2.dead","cum.stage3")
    ret<-data.frame(ret)
    ret<-ret[ret$day %in% tm,]
    return(ret)
}
```

S5 Text

The deterministic Diff model (R script). The output of the function is a data.frame of weekly model output starting from day 1, the day after initiation. The output comprise the following in the given order:

day	Day of observation (weekly)
stage2	Viable number of individuals in stage 2 at the time of observation
cum.stage2	Total number of individuals in stage 2 at the time of observation
cum.stage3	Total number of individuals in stage 3 at the time of observation

```
fun.Diff <- function(pr,tm,init) {
    stage1 <- init
    stage2 <- 0
    cum.stage2 <- 0
    cum.stage3 <- 0
    ret <- NULL
    p1 <- 1-pr[1]
    p3 <- 1-pr[3]
    pd <- p1^pr[2]
    PE <- (p1 - pd)/(1-pd)
    pl <- p3^pr[4]
    PL <- (p3 - pl)/(1-pl)
    GE <- (1-p1)*pd/(1-pd)
    GL <- (1-p3)*pl/(1-pl)
    for (n in 1:max(tm)) {
        new.stage2 <- GE * stage1
        new.stage3 <- GL * stage2
        cum.stage3 <- cum.stage3 + new.stage3
        cum.stage2 <- cum.stage2 + new.stage2
        stage2 <- new.stage2 + PL * stage2
        stage1 <- PE * stage1
        if (n %in% tm) {
            ret <- rbind(ret,c(n,stage2,cum.stage2,cum.stage3))
        }
    }
    colnames(ret) <- c("day","stage2","cum.stage2","cum.stage3")
    ret<-data.frame(ret)
    return(ret)
}
```

S6 Text

The deterministic ODE model (R script). The output of the function is a data.frame of weekly model output starting from day 1, the day after initiation. The output comprise the following in the given order:

day	Day of observation (weekly)
stage2	Viable number of individuals in stage 2 at the time of observation
cum.stage2	Total number of individuals in stage 2 at the time of observation
cum.stage3	Total number of individuals in stage 3 at the time of observation

```
fun.ODE <- function(pr,tm,init) {
    require(deSolve)
    funDet <- function(t, y, pr) {
        new.stage2 <- pr[2] * y[1]
        new.stage3 <- pr[4] * y[2]
        dStage1 <- - new.stage2 - pr[1] * y[1]
        dStage2 <- new.stage2 - new.stage3 - pr[3] * y[2]
        dCum.stage2 <- new.stage2
        dCum.stage3 <- new.stage3
        return(list(c(dStage1, dStage2, dCum.stage2, dCum.stage3)))
    }
    ret <- data.frame(ode(y = c(init,0,0,0),
                          times = tm,
                          func = funDet,
                          parms = pr)[,c(1,3,4,5)])
    colnames(ret) <- c("day","stage2","cum.stage2","cum.stage3")
    return(ret)
}
```

S7 Text

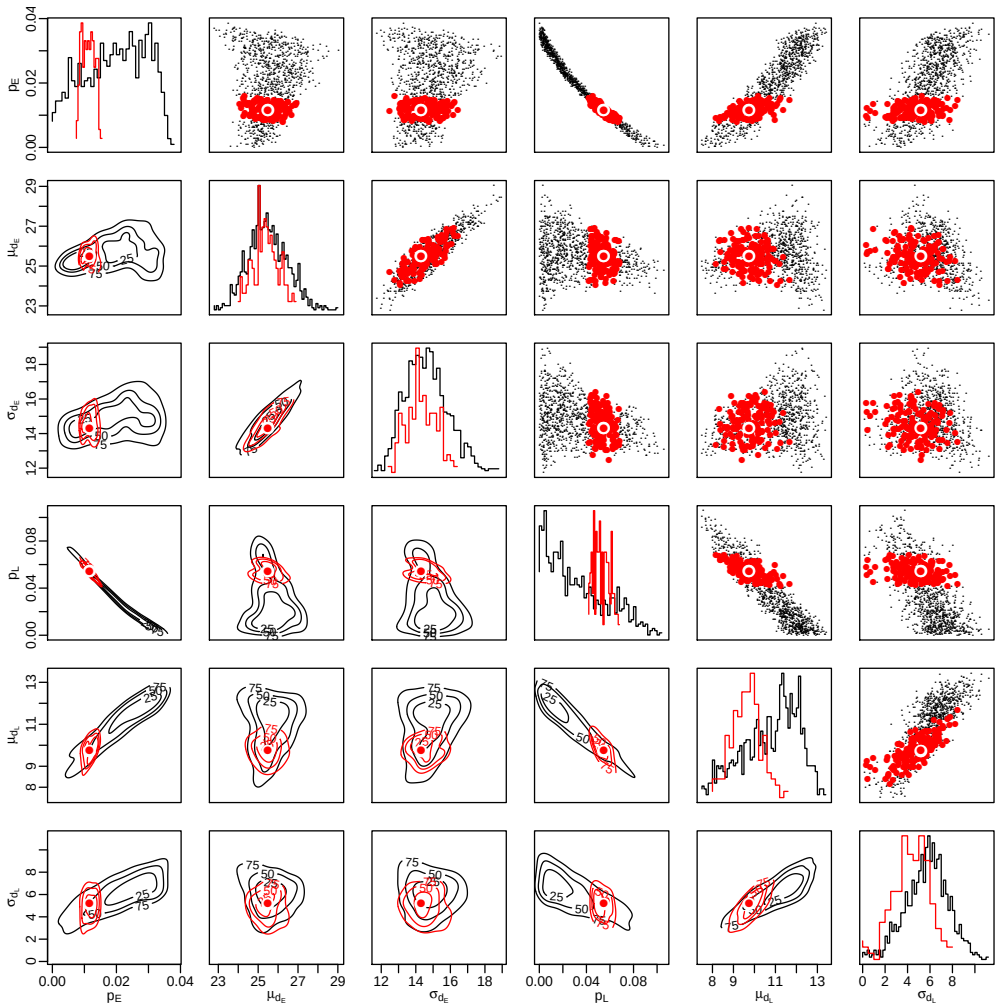
The deterministic DDE model (R script). The output of the function is a data.frame of weekly model output starting from day 1, the day after initiation. The output comprise the following in the given order:

day	Day of observation (weekly)
stage2	Viable number of individuals in stage 2 at the time of observation
cum.stage2	Total number of individuals in stage 2 at the time of observation
cum.stage3	Total number of individuals in stage 3 at the time of observation

```
fun.DDE <- function(pr,tm,init) {
    require(stagePop)
    ccFunctions <- list(
        deathFunc=function(stage,x,time,species,strain){
            if (species==1) return(c(pr[1],0)[stage])
            return(c(pr[1],pr[3],0)[stage])
        },
        durationFunc=function(stage,x,time,species,strain){
            if (species==1) return(c(pr[2],Inf)[stage])
            return(c(pr[2],pr[4],Inf)[stage])
        },
        reproFunc=function(x,time,species,strain){return(0)},
        emigrationFunc=function(stage,x,time,species,strain){return(0)},
        immigrationFunc=function(stage,x,time,species,strain){
            v <- 0
            if (stage==1 && pr[5]>0 && time>=0 && time<=pr[5]) {
                v <- init/pr[5]
            }
            return(v)
        }
    )
    out <- popModel(
        numSpecies = 2,
        numStages = c(2,3),
        timeDependLoss = c(TRUE,TRUE),
        timeDependDuration = c(FALSE,FALSE),
        ICs = list(matrix(0,ncol=1,nrow=2),matrix(0,ncol=1,nrow=3)),
        timeVec = seq(0,max(tm),0.1),
        rateFunctions = ccFunctions,
        stageNames = list(c("Stage1","Cum.stage2"),
                          c("Stage1","Stage2","Cum.stage3")),
        speciesNames = c("Vector1","Vector2"),
        saveFig = FALSE,
        plotFigs = FALSE
    )
    ret <- data.frame(out[out[,1] %in% tm,
                      c("time",
                        "Vector2.Stage2",
                        "Vector1.Cum.stage2",
                        "Vector2.Cum.stage3")])
    colnames(ret) <- c("day","stage2","cum.stage2","cum.stage3")
    return(ret)
}
```

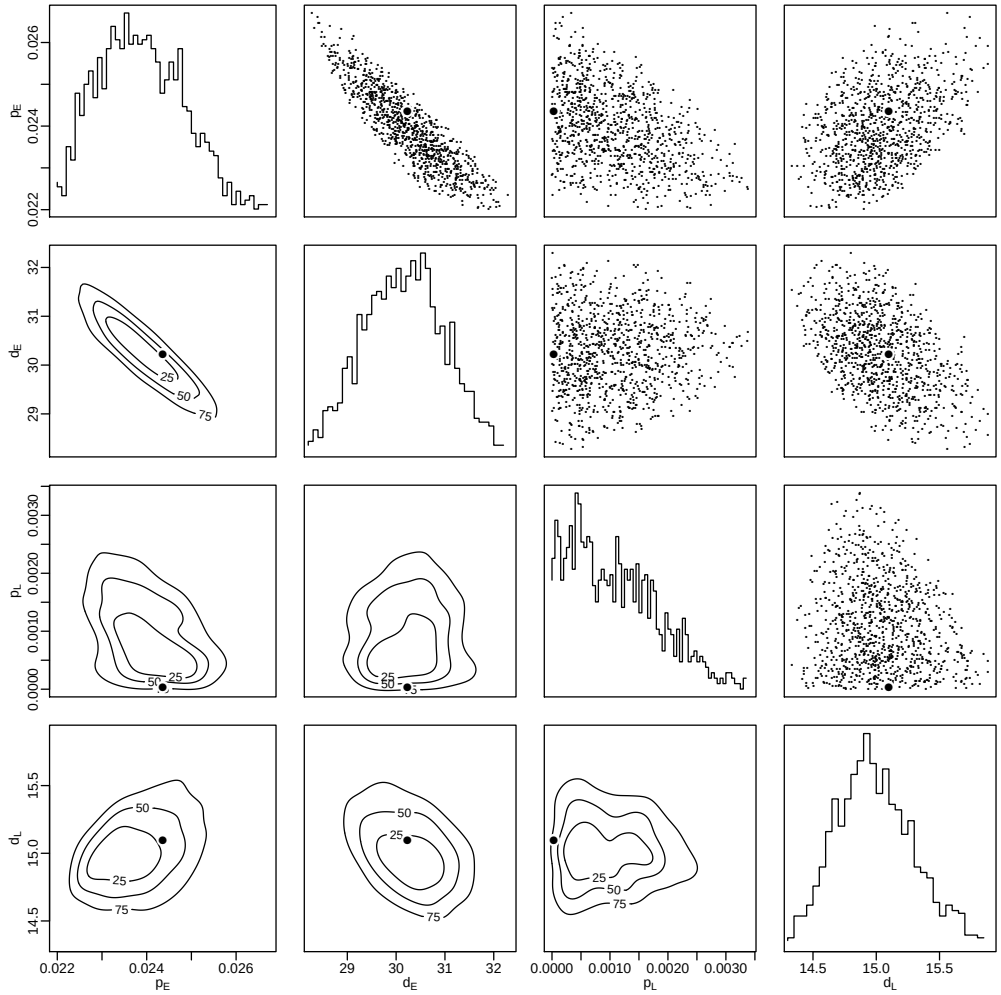
S1 Figure

The posterior samples for the sPop model. The 1000 parameter configurations inferred without using the number of dead larvae are plotted in black, and the 150 of them inferred using the number of dead larvae are plotted in red. The parameter configurations are plotted in pairs in the figures above the diagonal. Below the diagonal, kernel density estimates are given. Marginal frequency distributions for each parameter are given along the diagonal. The parameter configuration with the minimum distance is marked with a solid circle.



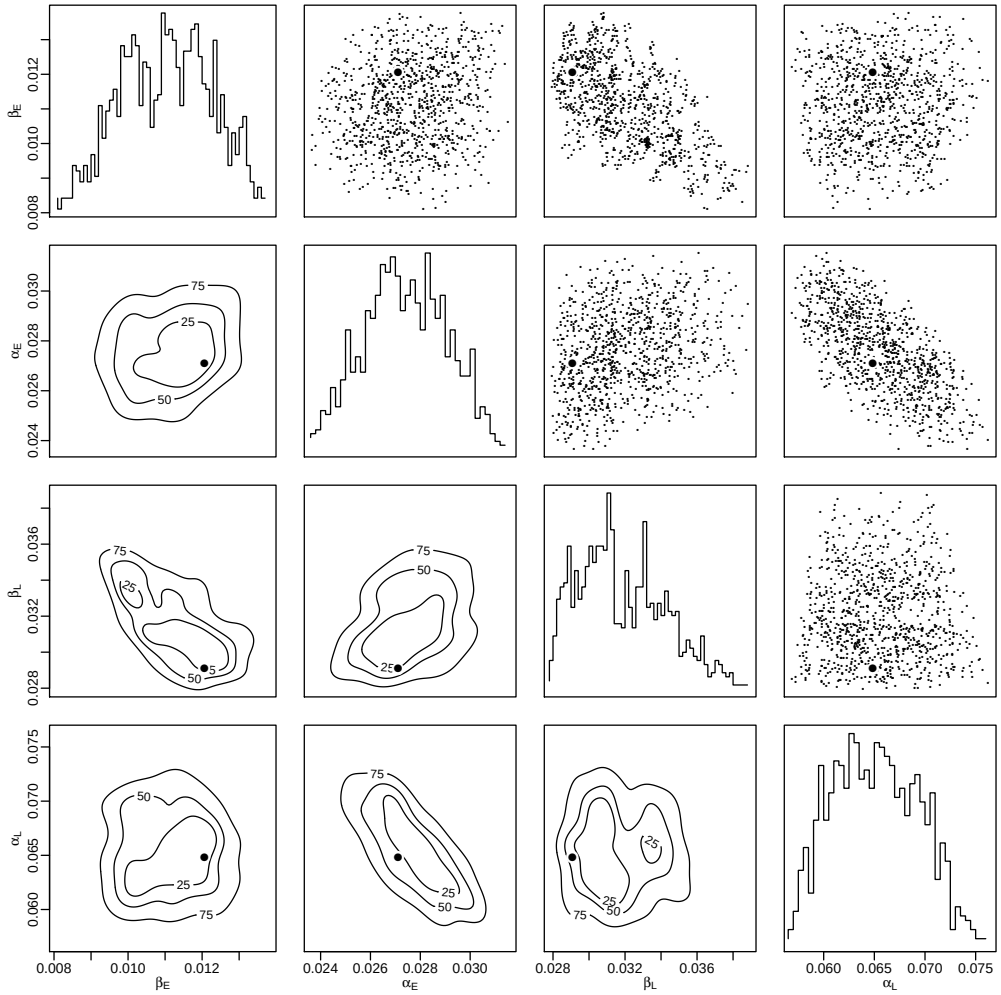
S2 Figure

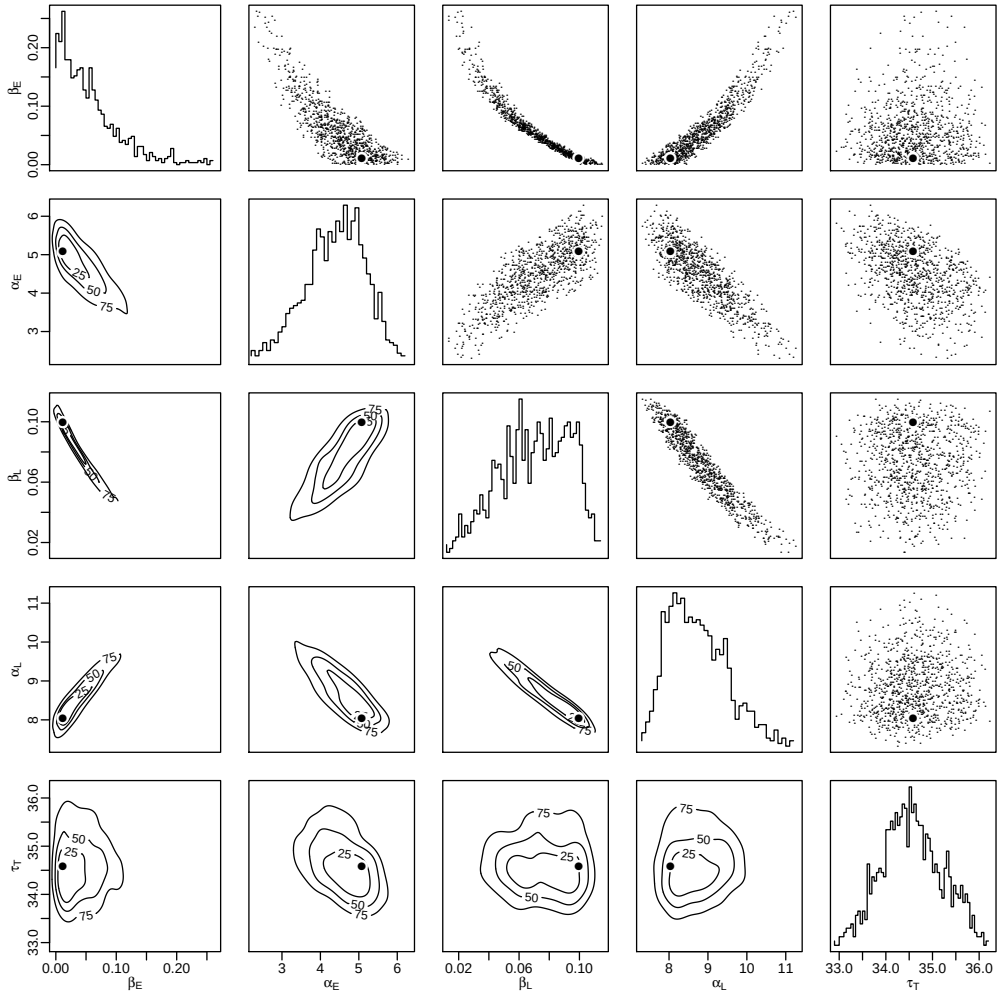
The posterior samples for the Diff model. 1000 parameter configurations are plotted in pairs in the figures above the diagonal. Below the diagonal, kernel density estimates are given. Marginal frequency distributions for each parameter are given along the diagonal. The parameter configuration with the minimum distance is marked with a solid circle.



S3 Figure

The posterior samples for the ODE model. 1000 parameter configurations are plotted in pairs in the figures above the diagonal. Below the diagonal, kernel density estimates are given. Marginal frequency distributions for each parameter are given along the diagonal. The parameter configuration with the minimum distance is marked with a solid circle.



S4 Figure

The posterior samples for the DDE model. 1000 parameter configurations are plotted in pairs in the figures above the diagonal. Below the diagonal, kernel density estimates are given. Marginal frequency distributions for each parameter are given along the diagonal. The parameter configuration with the minimum distance is marked with a solid circle.