

Projection of parasite numbers after fish mortality

When a host dies at time $t \geq 1$, it is often observed in most biological systems that the entire parasite population on the host becomes extinct before time $t + 1$. However, a study on survival and behaviour of *Gyrodactylus turnbulli* and *G. bullatarudis* on dead fish revealed that under certain conditions such as high parasite abundance at the time of death, these parasites might survive on a dead host for a while [1]. Thus, after premature host mortality (before the end of the infection period), we cannot assume that there is parasite population extinction at all body regions depending on the parasite abundance at the time of death. Consequently, during such realisations when an infected fish dies before the end of the observation period in the complex simulation model (developed under section 2 of the main paper) or observed empirical data, we use an estimated linear regression function (given by [equation 5](#)) based on the parasite data before host mortality, to linearly project the parasite numbers till the end of the observation period to aid in ABC implementation (by assigning more weight to the most recent data before host death as in a case of missing value imputation assuming at least missing-at-random). Furthermore, we minimise the projection estimation error by assigning more weight to the most recent outcomes or data prior to host mortality based on [equations 1–3](#). Thus, the parasite population projection after host mortality is used to aid in the summary statistics computation for ABC fitting of our complex stochastic model; specifically, during the calculation of other components of the summary statistics such as the log counts of parasites over time till day 17, and weights of the Wasserstein 1-D distance metric between body regions of the host. We propose a linear function to estimate missing values after fish mortality till the end of the observation period:

Let y_t = total parasites at time t , $z_t = \log(y_t)$, k be the time before fish mortality, α a regression parameter and $\gamma \in (0, 1)$, a tuning parameter. Given $z_1, z_2, z_3, \dots, z_k$, we want to estimate or project for $\hat{z}_{k+1}, \dots, \hat{z}_9$. Now, let the proposed least squares regression equation, denoted by $m(t)$, be defined as given by [equation 1](#);

$$m(t) = (k - t)\alpha + z_k \quad (1)$$

with the estimate of the regression parameter (α) determined such that

$$\hat{\alpha} = \min_t \sum (m(t) - z_t)^2 \gamma^{k-t}. \quad (2)$$

Thus, the prediction of z_i for $i = k + 1, k + 2, \dots, 9$ can be estimated using [equation 3](#)

$$\hat{z}_{k+i} = m(k+i) = -i \times \hat{\alpha} + z_k, \quad (3)$$

where i is the number of predictions across time. Then,

$$y_{k+i} = \exp(\hat{z}_{k+i}). \quad (4)$$

Hence, also letting $y_{t,u}$ be parasite at location u and time t implies the expected projections are defined such that

$$\hat{y}_{k+i,u} = \hat{y}_{k+i} \times \frac{y_{k,u}}{y_k}. \quad (5)$$

Remark. The estimate of the regression parameter (α) defined by [equation 2](#) can be derived explicitly using [Theorem 1](#) (for the theorem and its mathematical proof, see Appendix A1). Due to the high computational demand in simulating data from our complex stochastic model as well as implementing our weighted-iterative ABC for model fitting (especially during the computation of the high-dimensional summary statistics), we tuned or set the value of γ at ≥ 0.9 (based on all observed fish that prematurely died), which we found to be a reasonable choice based on some heuristic technique which achieved some degree of convergence when estimating α (defined by [Theorem 1](#) under Appendix A1) at varying γ values ranging from 0 to 1 across all the different parasite-host groups (see Appendix Figure A1 for a graphical summary, where each colour indicates an observed dead fish).

Appendix A1

Derivation of an exact least squares estimator of the regression parameter α (defined by [equation 2](#)) based on [Theorem 1](#).

Theorem 1. (Least squares estimate of the regression parameter)

The exact least squares estimate of the regression parameter α based on the linear regression [equation 1](#) is given as:

$$\hat{\alpha} = \frac{\sum_{t=1}^{k-1} (z_t - z_k)(k-t)\gamma^{k-t}}{\sum_{t=1}^{k-1} (k-t)^2 \gamma^{k-t}}. \quad (6)$$

Proof of [Theorem 1](#) .

Suppose $m(t) = (k-t)\alpha + z_k$, $\alpha = \sum_t (m(t) - z_t)^2 \gamma^{k-t}$ whilst fixing γ , and the loss function

$$L(\alpha, \gamma) = \min_{\alpha} \sum_t (m(t) - z_t)^2 \gamma^{k-t}. \quad (7)$$

Differentiating [equation 7](#) with respect to α and setting to zero gives

$$\frac{\partial L}{\partial \alpha} = 2 \sum_{t=1}^{k-1} \{(k-t)\alpha + z_k - z_t\} \gamma^{k-t} (k-t) = 0.$$

This implies

$$\sum_{t=1}^{k-1} (k-t)^2 \alpha \gamma^{k-t} + \sum_{t=1}^{k-1} (z_k - z_t) \alpha \gamma^{k-t} (k-t) = 0,$$

where

$$\alpha \sum_{t=1}^{k-1} (k-t)^2 \gamma^{k-t} = - \sum_{t=1}^{k-1} (z_k - z_t) \alpha \gamma^{k-t} (k-t).$$

Therefore, solving for α gives the required [equation 6](#) such that

$$\hat{\alpha} = - \frac{\sum_{t=1}^{k-1} (z_k - z_t) (k-t) \gamma^{k-t}}{\sum_{t=1}^{k-1} (k-t)^2 \gamma^{k-t}} = \frac{\sum_{t=1}^{k-1} (z_t - z_k) (k-t) \gamma^{k-t}}{\sum_{t=1}^{k-1} (k-t)^2 \gamma^{k-t}}, \gamma \in (0, 1). \quad \text{Q. E. D.}$$

Appendix Figure A1

A Figure showing the convergence effect of varying values of the tuning parameter $\gamma \in (0, 1)$ on the estimate of the regression coefficient α (defined by [equation 6](#)) across all parasite-fish groups (based on all observed fish that prematurely died as indicated by different colours). Appendix Figure A1 can be viewed and downloaded via this URL link: [click here](#).

Bibliography

1. Cable, J. and Harris, P. (2002). Gyrodactylid developmental biology: historical review, current status and future trends. *International Journal for Parasitology*, 32(3):255–280.