UNDER REVIEW

RESCIENCE C

Replication / Ecology

[Re] Modeling Insect Phenology Using Ordinal Regression and Continuation Ratio Models

Philipp H. Boersch-Supan[1,2]

[1]British Trust for Ornithology, Thetford, United Kingdom – [2]University of Florida, Gainesville, FL, USA

Edited by
(Editor)

Received
–

Published
–

DOI
–

1 Introduction

Phenology, the timing of seasonal biological phenomena, is a key aspect of plant and animal life. It defines the timing and duration of growth and reproduction and thereby determines the ability to capture seasonally variable resources [1]. Phenological analyses often focus on the timing of particular events, such as the dates of peak caterpillar abundance [2]. However, for many biological phenomena exact dates of particular events are more difficult to observe than the state of the system itself. For example, repeated but sparse survey visits may record whether a plant is in bud, flowering, or setting fruit, but not the exact dates when each of those stages was reached. Such observations can be used to categorize an organism's state into discrete classes which usually follow natural ordering, e.g. from least to most developed. The resulting data can be described using ordinal regression models [3, 4] which then allow inferences about phenological progression.

I here replicate a number of ordinal regression models that were employed by Dennis, Kemp, and Beckwith[5] and Candy[6] to describe insect phenology.

2 Data

The models replicated in this study are fitted to a data set on the phenology of the western spruce budworm *Choristoneura freemani* (Lepidoptera: Tortricidae), a defoliating moth that is widespread in western North America [7]. This data set was originally published in [5] and is a subset of a larger budworm survey data set analysed in [8]. The data consist of 12 sampling occasions at which counts of individual budworms in each of seven development stages (five larval instars, pupae, and adults) were recorded. The only available covariate is a measure of seasonal progression, the accumulated degree days calculated using a threshold of 5.5°C. Candy[6] noted an inconsistency in these data, namely that the reported total number of individuals did not correspond to the sum across the seven development stages for two of the sampling occasions. I therefore use the data set as it was republished in [9], where numbers in each stage have been assumed correct and the totals for each sampling occasion were adjusted accordingly.

Correspondence should be addressed to Philipp H. Boersch-Supan (pboesu@gmail.com)
The authors have declared that no competing interests exist.
Code is available at https://github.com/pboesu/replication_candy_1991..
Data is available at https://github.com/pboesu/replication_candy_1991.

Table 1. The data used in this replication are counts of western spruce budworm *Choristoneura freemani* across seven developmental stages on 12 sampling occasions, as published in [9].

Degree Days	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
58	16	0	0	0	0	0	0
82	10	0	0	0	0	0	0
107	23	7	0	0	0	0	0
155	3	44	0	0	0	0	0
237	0	6	45	13	0	0	0
307	0	2	9	48	15	0	0
342	0	0	1	34	37	0	0
388	0	0	1	10	87	5	0
442	0	0	0	7	53	21	0
518	0	0	0	0	10	65	1
609	0	0	0	0	0	14	26
685	0	0	0	0	0	0	42

3 Methods

The statistical models replicated here are different types of ordinal regression models [4], all with the aim of predicting the proportion of an insect population in a particular development stage at any given time. In particular, they represent three different parameterisations of the so-called cumulative model and one version of the so-called sequential model. A summary of the theory underlying these models and their derivation is provided in [10].

Both model types assume that the development of an insect follows an unobservable stochastic process $S(t)$ made up of accumulated increments of development over time t. As $S(t)$ increases, the insect passes through successive stages $j = 1, \ldots, r$, delimited by $r - 1$ moults, with the jth moult occurring when the development threshold a_j is reached:

$$\begin{aligned}
\text{stage } 1: &\quad S(t) \leq a_1 \\
\text{stage } 2: &\quad a_1 < S(t) \leq a_2 \\
&\vdots \qquad\qquad \vdots \\
\text{stage } r-1: &\quad a_{r-2} < S(t) \leq a_{r-1} \\
\text{stage } r: &\quad a_{r-1} < S(t)
\end{aligned}$$

The a_j values are typically unknown and their estimation from observed data was the goal of the orginal studies [5, 6].

3.1 Cumulative model with constant variance

The ordinal regression model of [6] is specified in terms of the cumulative number of individuals $m_{ij} = \sum_{k=1}^{j} n_{ik}$ observed in stages 1 to j on a sampling occasion i.

$$\begin{aligned}
\mathbf{E}(m_{ij}) &= N_i Pr(S(t) < \alpha_j), \qquad j = 1, \ldots, r &(1)\\
&= N_i G(\alpha_j + \beta z_i) &(2)
\end{aligned}$$

where G is the cumulative probability density function of $S(t)$, α_j are ordered thresholds or cut-point parameters, β is a vector of regression parameters and z_i is a vector of predictor variables. If the probability of an individual being in stage j or earlier at time t_i is

$$\mu_{ij} = \mathbf{E}(m_{ij})/N_i$$

one can interpret G^{-1} as the link function of a generalised linear model (GLM) with the linear predictor

$$\eta_{ij} = \alpha + \beta z_i$$

This ordinal regression model is commonly known as the cumulative model [10], and is applied to the budworm data in [6] using the logit and complementary log-log (cloglog) link functions. In both cases the parameterisation results in a constant variance for $S(t)$. For the purpose of parameter estimation Candy[6] re-expressed the model in terms of stage-specific counts n_{ij}

$$\mathbf{E}(n_{ij}) = N_i\{G(\alpha_j + \beta z_i) - G(\alpha_{j-1} + \beta z_i)\} \tag{3}$$

and fits it using a Poisson likelihood [11]. No code or initial values for the likelihood optimisation are provided for this estimation procedure in the original paper. I therefore created an R version of the estimation procedure which directly optimizes a Poisson log-likelihood for Equation 3 using the BFGS method in the `optim` function. Initial values for the optimisation were determined using a random search across plausible start values (see Appendix A.1). The cumulative model is implemented in various R packages, including in the `vglm` function in `VGAM` [12] and the `clm` function in `ordinal` [13] and for comparison I also attempted to fit the models using these functions using their default settings.

3.2 Cumulative model with proportional variance

Dennis, Kemp, and Beckwith[5] used a different parameterisation of the ordinal model with a logit link, i.e. assuming a logistic distribution for $S(t)$, such that the probability that an insect's development at time t has not exceeded s amounts to

$$Pr(S(t) \leq s) = \left\{1 + \exp\left[-\left(\frac{s-t}{\sqrt{b^2 t}}\right)\right]\right\}^{-1} \tag{4}$$

where b^2 is a positive constant. The cumulative distribution function in Eqn. 4 corresponds to a logistic distribution with a mean of t and a variance which increases proportional to the mean as $(\pi^2/3)b^2t$. At any fixed time t the thresholds a_j segment the probability distribution function in to r parts and the area under the curve between a_{j-1} and a_j gives the probability that the insect will be in stage j at time t.
This modelling approach is applied to data consisting of samples i that record the number of insects x_{ij} in stage j at times $t_1, t_2, \ldots, t_q$ and the x_{ij} are assumed to be random samples from a multinomial distribution with corresponding multinomial probabilities p_{ij}

$$p_{ij} = Pr(a_{j-1} < S(t_i) \leq a_j) \tag{5}$$

$$= \left\{1 + \exp\left[-\left(\frac{a_j - t_i}{\sqrt{b^2 t_i}}\right)\right]\right\}^{-1} - \left\{1 + \exp\left[-\left(\frac{a_{j-1} - t_i}{\sqrt{b^2 t_i}}\right)\right]\right\}^{-1} \tag{6}$$

To fulfill the constraint that $\sum_{j=1}^{r} p_{ij} = 1$, it is further assumed that $a_0 = -\infty$ and $a_r = +\infty$. The model has r unknown parameters $a_1, \ldots, a_{r-1}$ and b^2 which can be found by maximising the corresponding log-likelihood function which takes the form

$$\ell = \log C + \sum_{j=1}^{r}\sum_{i=1}^{q} x_{ij} \log p_{ij} \tag{7}$$

where C is a combinatorial constant that is independent of the parameter values.
Dennis, Kemp, and Beckwith[5] provided SAS code and initial values to estimate the parameters under this likelihood using an interatively reweighted non-linear least squares approach based on `PROC NLIN`. This was updated to run in a contemporary version of SAS (SAS 9.4) and is provided in the article code repository. However, since SAS is a proprietary software package, I implemented two R versions of the estimation procedure which both directly optimize the log-likelihood (7) using the L-BFGS-B optimisation routine [14] in the R `optim` function. The first implemetantion is a direct translation of the SAS code which uses a literal implementation of the logistic cumulative distribution function and employs a lower parameter bound of 0 and initial values provided in [5] for the estimation. The second implementation makes use of the R function `stats::plogis` to implement the logistic CDF and employs a lower parameter bound of 10^{-12} and the same initial values as above.
Candy[6] re-expressed the proportional variance model (Eqn. 6) to match the form of Eqn. 3, which results in the following reparameterisation $\alpha_j = a_j/b$, $\beta = -1/b$, and $z_i = \sqrt{t_i}$, and uses the Poisson likelihood approach described in section 3.1 for parameter estimation. A set of example macros for the software package `GLIM` [15] is provided in an earlier manuscript by the same author [9]. GLIM is no longer actively developed or distributed, but initial values from the GLIM code were used in the estimation with the BFGS method of R `optim`.

3.3 Sequential model

A different class of ordinal regression model can be derived by treating the observations as the result of a strictly ordered counting process, i.e. to achieve a stage j, all lower stages $1, \ldots, j-1$ have to be achieved first. The general form of this model is known as the sequential model, and rather than assuming a single latent process $S(t)$ as in the cumulative model there is a latent continuous variable S_j for each category j [10]. Analogous to the cumulative model it can be framed as a GLM

$$S_j = \eta + \epsilon_j \tag{8}$$

with a linear predictor η and an error term ϵ_j which has mean zero and is distributed following some distribution G. This leads to a model of the form

$$Pr(S = a_j | S \geq a_j, \eta) = G(a_j - \eta) \tag{9}$$

When G is the logistic distribution this model is also known as the continuation ratio model [16, 10]. Confusingly, there are two common versions of the model in the literature both using this name. The one outlined above describing the probability of the sequential process *stopping* at stage j, and the other describing the probability of the process *continuing* beyond stage j, i.e.
$Pr(S > a_j | S \geq a_j, \eta)$ [10, 12].
The paper replicated here [6] used the stopping parameterisation of this model. In their notation the expected value for the stage-specific counts n_{ij} is

$$\begin{aligned} \mathbf{E}(n_{ij}) &= N_i G(\beta_{01} + \beta_{11} t_i), \qquad j = 1 \\ &= N^*_{ij} G(\beta_{0j} + \beta_{1j} t_i), \qquad j = 2, \ldots, r-1 \end{aligned} \tag{10}$$

with

$$N^*_{ij} = \left(N_i - \sum_{k=1}^{j-1} n_{ik} \right) \tag{11}$$

and conditional probabilities

$$p^*_{ij} = G(\beta_{0j} + \beta_{1j} t_i), \qquad j = 1, \ldots, r-1 \tag{12}$$

Table 2. Parameter estimates for the cumulative model with constant variance (Eqn. 3). This table replicates results presented in the first two rows of Table 2 of [6]. Note that `ordinal::clm` uses a parameterisation $\alpha_j - \beta z_i$ for the linear predictor yielding a parameter estimate for β with the opposite sign than the other methods. The cloglog link model failed to fit using `VGAM::vglm`.

Method	α_1	α_2	α_3	α_4	α_5	α_6	β	Link
Original [6]	5.49	9.39	12.23	15.70	21.26	27.25	-0.05	logit
R `optim`	5.47	9.38	12.22	15.69	21.25	27.22	-0.05	logit
R `clm`	5.47	9.36	12.21	15.67	21.22	27.19	0.05	logit
R `vglm`	5.47	9.36	12.21	15.67	21.22	27.19	-0.05	logit
Original [6]	3.32	5.85	7.72	10.03	13.46	17.52	-0.03	cloglog
R `optim`	3.35	5.90	7.79	10.12	13.57	17.67	-0.03	cloglog
R `clm`	3.32	5.85	7.71	10.02	13.46	17.52	0.03	cloglog
R `vglm`	NA	NA	NA	NA	NA	NA	NA	cloglog

Table 3. Parameter estimates for the cumulative logit model with proportional variance. This table replicates results presented in the first row of Table 1 of [8] and the last row of Table 2 of [6].

Method	a_1	a_2	a_3	a_4	a_5	a_6	b^2	Eqn.
Original [8]	121.080	204.360	264.410	342.473	465.620	599.570	1.559	6
SAS `NLIN`	120.000	204.700	264.600	341.300	464.500	595.700	1.412	6
R `optim`	120.044	204.658	264.591	341.296	464.470	595.704	1.412	6
R `optim/plogis`	120.038	204.671	264.590	341.292	464.473	595.705	1.412	6
Method	α_1	α_2	α_3	α_4	α_5	α_6	β	Eqn.
Original [6]	101.000	172.200	222.700	287.200	390.900	501.300	-0.842	3
R `optim`	100.989	172.181	222.598	287.134	390.771	501.157	-0.841	3

Candy[6] uses GLM estimation routines in `GLIM` to fit the $r-1$ models defined in Eqn. 10 assuming the n_{ij} are binomially distributed conditional on N_i for stage 1 and conditional on N^*_{ij} for stages $2, \ldots, r-1$. No code is provided for this estimation procedure in the original paper, but the model is straightforward to implement using the `glm` function in R with a model formula of the form

```
cbind(count,total - N_star) ~ stage + stage:time - 1
```

where the variable `count` represents the n_{ij}, `N_star` represents N_i for $j = 1$ and N^*_{ij} for all other observations, `stage` is a factor variable encoding j and `time` are the t_i. The sequential model with stopping parameterisation is also implemented in `VGAM::vglm` [12] and for comparison I attempted to fit the model using this function with its default settings.

4 Results

4.1 Cumulative model with constant variance

Parameters were estimated using a direct optimisation of the Poisson likelihood for Eqn. 3, as well as with the R functions `VGAM::vglm` and `ordinal::clm`. The cloglog link model failed due to numerical errors when using the `vglm` function. Parameter estimates were close to those of the original study for the two R packages, and differed slightly for the `optim` method (Table 2), the latter exhibiting a noticeable sensitivity to the choice of starting values (see Appendix).

4.2 Cumulative model with proportional variance

The original SAS code provided in [5] required minimal updates to run in a contemporary version of SAS (SAS 9.4). Translating the model code to R was straightforward once I took the decision to implement a direct minimisation of the negative log likelihood with `optim`.

Parameter estimates from SAS `NLIN` and both R `optim` implementations (Table 3) were virtually identical, but differed slightly from the parameter estimates presented in [8], which was assumed to be the original source for the parameter estimates, as no parameter estimates were presented in [5]. The observed differences in estimates were $< 1\%$ for the cut-point parameters α_i, but c. 9% for β.

Based on these three sets of parameter estimates it was also possible to redraw two figures from [5]. Figure 1 and 2, respectively, show that despite the observed parameter differences there is an overall good agreement between the original results and the replication.

Table 4. Parameter estimates for the sequential model with stopping ratios (Equation 10). This table replicates results presented in Table 3 of [6]. The cloglog link model failed to fit using `VGAM::vglm`.

Method	Parameter	β_1	β_2	β_3	β_4	β_5	β_6	Link
Original [6]	β_{0j}	10.410	12.960	12.020	11.160	17.700	33.730	logit
R `glm`	β_{0j}	10.410	12.959	12.020	11.165	17.698	33.726	logit
R `vglm`	β_{0j}	10.410	12.959	12.020	11.165	17.698	33.726	logit
Original [6]	β_{1j}	-0.085	-0.062	-0.046	-0.033	-0.038	-0.056	logit
R `glm`	β_{1j}	-0.085	-0.062	-0.046	-0.033	-0.038	-0.056	logit
R `vglm`	β_{1j}	-0.085	-0.062	-0.046	-0.033	-0.038	-0.056	logit
Original [6]	β_{0j}	7.350	8.530	9.120	8.440	10.090	16.300	cloglog
R `glm`	β_{0j}	7.347	8.537	9.124	8.442	10.087	16.298	cloglog
R `vglm`	β_{0j}	NA	NA	NA	NA	NA	NA	cloglog
Original [6]	β_{1j}	-0.065	-0.044	-0.037	-0.026	-0.023	-0.029	cloglog
R `glm`	β_{1j}	-0.065	-0.044	-0.037	-0.026	-0.023	-0.029	cloglog
R `vglm`	β_{1j}	NA	NA	NA	NA	NA	NA	cloglog

4.3 Sequential model

Parameters were estimated using the R `glm` function and `VGAM::vglm`. The GLM formulation of the model fitted with R `glm` produced a warning that fitted probabilities numerically 0 or 1 occurred. This is because the population variation in development speed for the study species is low compared to the overall speed of seasonal progression. As a result the early and late development stages of the observed species never occur together, and the observed data are sparse in the sense that many cell counts in Table 1 that are 0. As a consequence the model estimates large effects, which although estimable on the link scale, are indistinguishable from 1 on the probability scale given the limitations of floating point representations. The cloglog link model failed due to numerical errors when using the `vglm` function. Parameter estimates, where obtained, were identical to the original study (Table 4) within the precision reported in [6].

5 Discussion

Overall the results from both [5] and [6] could be replicated closely.

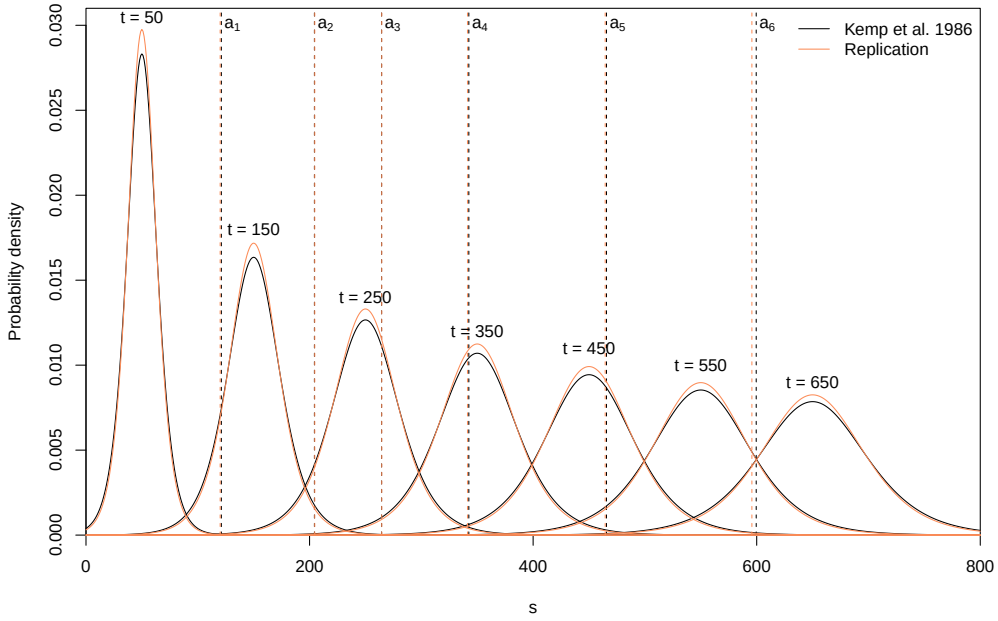

Figure 1. Logistic PDF of the cumulative model with proportional variance (Equation 6) plotted for fixed values of t. Area under the PDF between a_{j-1} and a_j gives the expected proportion of insects in stage j at time t. The graph is based on the estimates in Table 1 of [8] (black lines) and the estimates from the replication using the R `optim` implementation (red lines). This figure replicates Figure 2 in [5].

The SAS code provided in [5] required only minimal updates to run in a contemporary version of SAS (SAS 9.4) and produced virtually identical estimates as the R re-implementation. These estimates, however, differed slightly from the parameter estimates reported in [8]. Given that the same initial values were used in all implementations, I believe that this disagreement is most likely caused by the inconsistencies in the published data set described in Section 2. The corrections applied to the data by [6] result in a data set that is internally consistent but likely different to that on which the estimates in [8] are based.

No code was provided in [6]. However, the mathematical and verbal descriptions of the models were detailed enough to re-implement the estimation procedures in R. GLIM code of the cumulative model with proportional variance was available from an earlier manuscript [9]. This allowed me to use the same initial values as the original study for this model. Initial values for the model with constant variance had to be guessed. The direct optimisation of the likelihood is sensitive to the choice of initial values [5]. Additional simulations (Appendix A.1) showed that this was particularly the case for the cloglog-link model. The coefficient of variation for estimates derived from a wide range of initial values was c. 0.5% and c. 1.7% for the logit and cloglog models, respectively, whenenver the model converged succesfully (Table S1). The observed differences between the published parameter estimates and those obtained using R `optim` (Table 3) are well within this range of variation. Another factor may be slight differences in the numerical implementation of the inverse link functions. Naive implementations of both the inverse logit and inverse complementary log-log function suffer from numerical underflow and/or overflow. The GLIM code from [9] uses multiple thresholding steps during the calculation of the linear predictor to mitigate against this, whereas the R implementation makes use of a single thresholding step in the inverse link function (`gtools::inv.logit` [17] or `stats::plogis` and `VGAM::cloglogLink`, respectively). While differences in parameter estimates between a literal implementation of the logit link and `stats::plogis` for the cumulative model with proportional variance (Table 4) were less than 0.01%, the complementary log-log function proved to be numerically much less stable (Figure S1) and this may have contributed to the failure to fit either of the two corresponding models using the `VGAM::vglm` function. However, despite several troubleshooting attempts the exact reason for the numerical issues leading to the failure could not be established. The issue has been reported to the maintainer of the `VGAM` package as a potential bug.

References

1. I. Chuine and J. Régnière. "Process-Based Models of Phenology for Plants and Animals." In: **Annual Review of Ecology, Evolution, and Systematics** 48.1 (2017), pp. 159–182. DOI: 10.1146/annurev-ecolsys-110316-022706.
2. J. D. Shutt, M. D. Burgess, and A. B. Phillimore. "A Spatial Perspective on the Phenological Distribution of the Spring Woodland Caterpillar Peak." In: **The American Naturalist** 194.5 (2019), E109–E121. DOI: 10.1086/705241.
3. P. McCullagh. "Regression Models for Ordinal Data." In: **Journal of the Royal Statistical Society: Series B (Methodological)** 42.2 (Jan. 1980), pp. 109–127. DOI: 10.1111/j.2517-6161.1980.tb01109.x.
4. A. Agresti. **Analysis of Ordinal Categorical Data**. John Wiley & Sons, Inc., Mar. 2010. DOI: 10.1002/9780470594001.
5. B. Dennis, W. P. Kemp, and R. C. Beckwith. "Stochastic Model of Insect Phenology: Estimation and Testing." In: **Environmental Entomology** 15.3 (June 1986), pp. 540–546. DOI: 10.1093/ee/15.3.540.
6. S. G. Candy. "Modeling insect phenology using ordinal regression and continuation ratio models." In: **Environmental entomology** 20.1 (1991), pp. 190–195. DOI: 10.1093/ee/20.1.190.
7. M. H. Brookes, R. W. Campbell, J. J. Colbert, R. G. Mitchell, and R. W. Stark. **Western spruce budworm**. Cooperative State Research Service Technical Bulletin 1694. United States Department of Agriculture Forest Service, 1987.

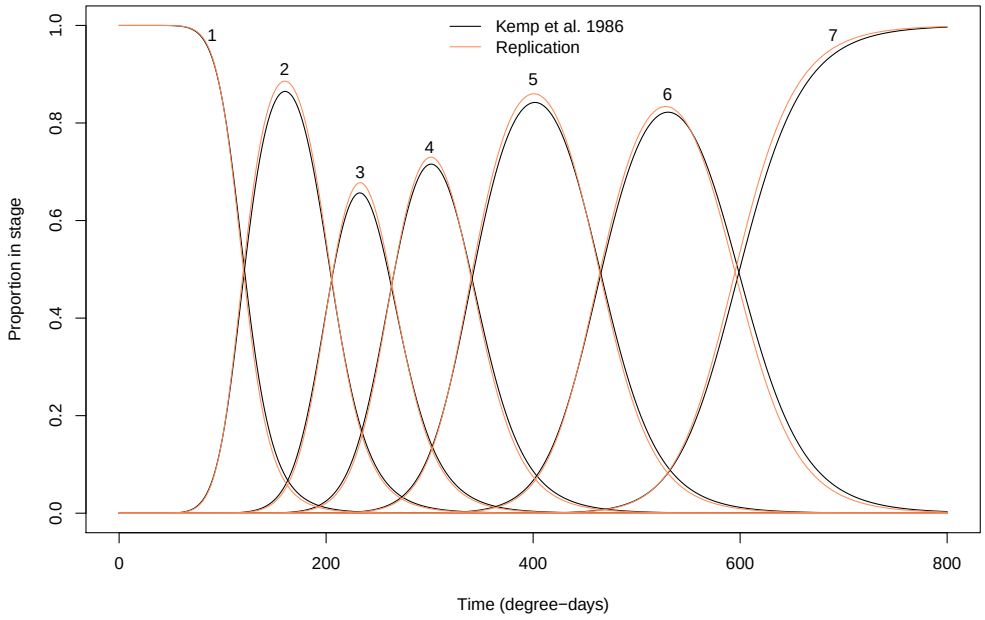

Figure 2. Proportion of insects expected in stages 1-7 under the cumulative logit model with proportional variance (Equation 6) plotted as functions of time t. Values of a_j and b^2 used in the graph are the estimates given in Table 1 of [8] (black lines) and the estimates from the replication using the R `optim` implementation (red lines). This figure replicates Figure 3 in [5].

8. W. P. Kemp, B. Dennis, and R. C. Beckwith. "Stochastic Phenology Model for the Western Spruce Budworm (Lepidoptera: Tortricidae)." In: **Environmental Entomology** 15.3 (June 1986), pp. 547–554. DOI: 10.1093/ee/15.3.547.
9. S. G. Candy. "Biology of the mountain pinhole borer, Platypus subgranosus Schedl, in Tasmania." MA thesis. University of Tasmania, 1990. URL: https://eprints.utas.edu.au/18864/.
10. P.-C. Bürkner and M. Vuorre. "Ordinal Regression Models in Psychology: A Tutorial." In: **Advances in Methods and Practices in Psychological Science** 2.1 (Feb. 2019), pp. 77–101. DOI: 10.1177/2515245918823199.
11. R. Thompson and R. Baker. "Composite link functions in generalized linear models." In: **Journal of the Royal Statistical Society: Series C (Applied Statistics)** 30.2 (1981), pp. 125–131. DOI: 10.2307/2346381.
12. T. W. Yee. "The VGAM Package for Categorical Data Analysis." In: **Journal of Statistical Software** 32.10 (2010), pp. 1–34. DOI: 10.18637/jss.v032.i10.
13. R. H. B. Christensen. **ordinal—Regression Models for Ordinal Data**. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal. 2019.
14. R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. "A limited memory algorithm for bound constrained optimization." In: **SIAM Journal on scientific computing** 16.5 (1995), pp. 1190–1208.
15. M. Aitkin. **Statistical modelling in GLIM**. Oxford Oxfordshire New York: Clarendon Press Oxford University Press, 1989.
16. S. E. Fienberg. "The analysis of cross-classified categorical data." In: **Massachusetts Institute of Technology Press, Cambridge and London** (1980).
17. G. R. Warnes, B. Bolker, and T. Lumley. **gtools: Various R Programming Tools**. R package version 3.8.2. 2020. URL: https://CRAN.R-project.org/package=gtools.

A Appendix: Sensitivity of the optimisation procedure to intial values

A.1 Approach

The sensitivity of parameter estimates for the cumulative model with constant variance (Eqn. 3) obtained using the direct likelihood optimisation with `optim` was assessed by simulation. Simulations consisted of drawing a set of random initial values. To preserve the ordering of the cut-point parameters α_i, their initial values α_i° were assembled as the cumulative sum of six independent draws

$$x_j \sim \text{Uniform}(1, 20),$$

i.e.

$$\alpha_i^\circ = \sum_{j=1}^{i} x_j.$$

The initial value for β was drawn as

$$\beta^\circ \sim \text{Uniform}(-1, 1).$$

Optimisation proceeded using the likelihood outlined in Section 3.1 with a numerical threshold to prevent the Poisson likelihood from numerical underflowing. As this allowed convergence at an infinite likelihood, the likelihood was evaluated one final time without thresholding at the converged parameter values and parameters were only retained when the unthresholded likelihood was finite. Parameters were further filtered to remove parameter sets where convergence was achieved at a finite likelihood, but with a substantially larger overall log-likelihood than the best model fit. Thresholds for this were determined empirically and set at 93 for the logit link and 100 for the cloglog model.

A.2 Results

The optimisation was reasonably robust for the logit link model. 82% of simulations achieved convergence and of these over 98% met the filtering threshold (Figure S1). Among the filtered parameter sets variability in the final estimates was small (parameter-wise coefficients of variation were all smaller than 0.5%) and there was no correlation between initial values and parameter estimates (Table S1). Convergence failed for initial values $\beta^\circ > 0.8$, and convergence failures were likely, but not guaranteed for $\beta^\circ \approx 0$ (Figure S2).

The optimisation of the cloglog model was much less stable. Only 54% of simulations resulted in convergence, and of those fewer than half (43%) met the filtering threshold. This is apparent as a large cluster of estimates deviating from the reference estimates in Figure S1. Variability in the filtered parameter estimates was larger than for the logit model, but still small in absolute terms (parameter-wise coefficients of variation were all between 1.6% and 2.0%) and there was no correlation between initial values and parameter estimates (Table S1). Convergence failed for many $0 < \beta^\circ < 0.2$, and always for $\beta^\circ > 0.2$ (Figure S3). No general pattern was apparent for initial value combinations that led to the model converging at local optima far from the optimum associated with the reference parameter values.

Table S1. Summary statistics of parameter sensitivities to starting values for the cumulative model with constant variance (Eqn. 3). The coefficients of variation (CV), correlation coefficients between initial and convergend values ρ°, and their corresponding p-values P_{ρ} were calculated for the filtered parameter estimates only.

Par.	Link	Sim. range	Conv. range	Filtered range	CV (%)	ρ^{o}	P_{ρ}
a1	logit	1.3 - 20	1.3 - 20	1.3 - 20	0.499	0.025	0.701
a2	logit	3.1 - 37.8	3.1 - 37.6	3.1 - 37.6	0.480	0.090	0.164
a3	logit	5.4 - 54.1	5.4 - 54.1	5.4 - 54.1	0.463	0.051	0.430
a4	logit	13 - 69.9	15.9 - 69.9	15.9 - 69.9	0.445	0.026	0.682
a5	logit	22.9 - 85	22.9 - 85	22.9 - 85	0.430	-0.015	0.819
a6	logit	31.6 - 100.7	32.7 - 100.7	32.7 - 100.7	0.448	-0.040	0.538
b	logit	-1 - 1	-1 - 0.8	-1 - 0.8	0.440	-0.064	0.320
a1	cloglog	1.1 - 19.9	1.1 - 19.9	1.2 - 18.7	1.986	0.034	0.778
a2	cloglog	3.3 - 38.9	4 - 37.3	4 - 34.4	1.774	-0.045	0.709
a3	cloglog	7.9 - 57	9.7 - 52.4	9.7 - 50.7	1.708	-0.025	0.837
a4	cloglog	15 - 67	16.3 - 66.9	16.3 - 63.8	1.689	-0.055	0.651
a5	cloglog	18.4 - 77.8	18.4 - 77.8	18.4 - 77.8	1.627	0.010	0.935
a6	cloglog	22.8 - 91.6	30.7 - 91.6	30.7 - 83.2	1.663	0.039	0.751
b	cloglog	-1 - 1	-1 - 0.2	-1 - 0.2	1.678	-0.034	0.779

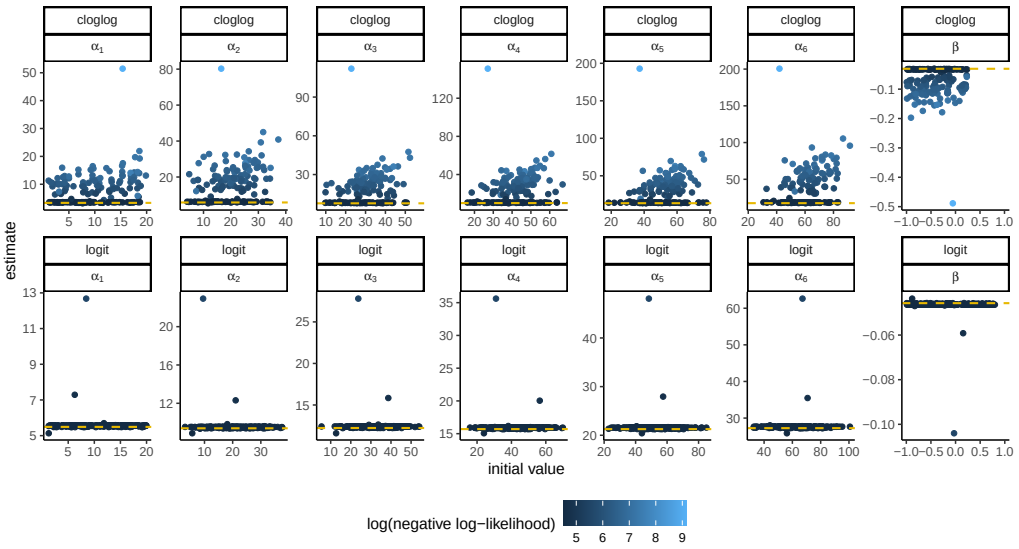

Figure S1. Optimisation results for randomly drawn initial values for the cumulative model with constant variance. Parameter estimates are coloured by the log of the negative log-likelihood of the corresponding model fit. Yellow dashed lines correspond to parameter values reported in [6]. The majority of initial values resulted in estimates close to the original publication for the logit model (bottom), but convergence to local optima far from the published parameters was common in the cloglog link model (top).

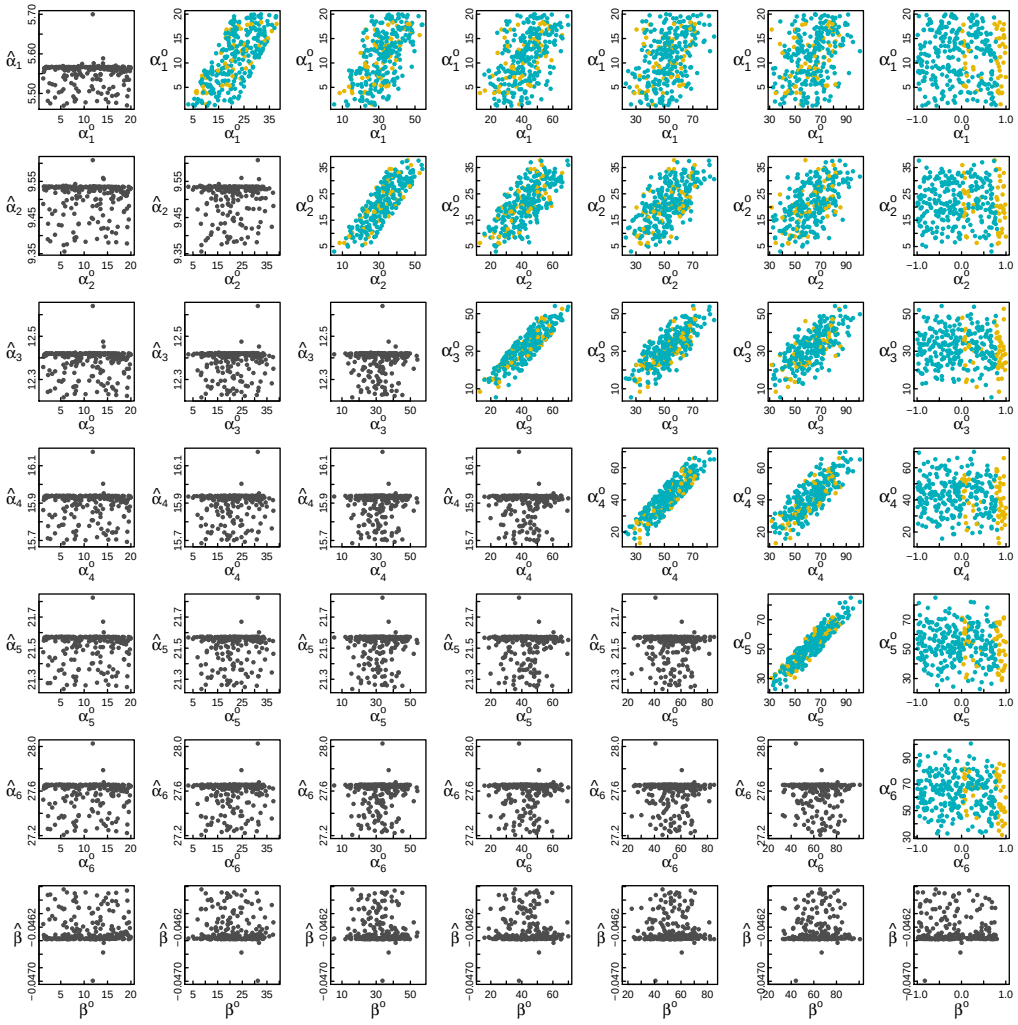

Figure S2. Sensitivity of optimisation results to initial values for the logit-link cumulative model with constant variance. The lower triangle and diagonal elements of the scatterplot matrix (grey symbols) show pairwise plots of filtered parameter estimates $\hat{\alpha}_i, \hat{\beta}$ against corresponding initial values $\alpha_i^\circ, \beta^\circ$. The upper triangle shows pairwise plots of the initial values for the same model parameters against each other. Green dots indicate succesfull convergence of the optimisation, yellow dots indicate convergence failures. Optimisation failed whenever $\beta^\circ > 0.8$, furthermore convergence failures were more likely when $\beta^\circ \approx 0$.

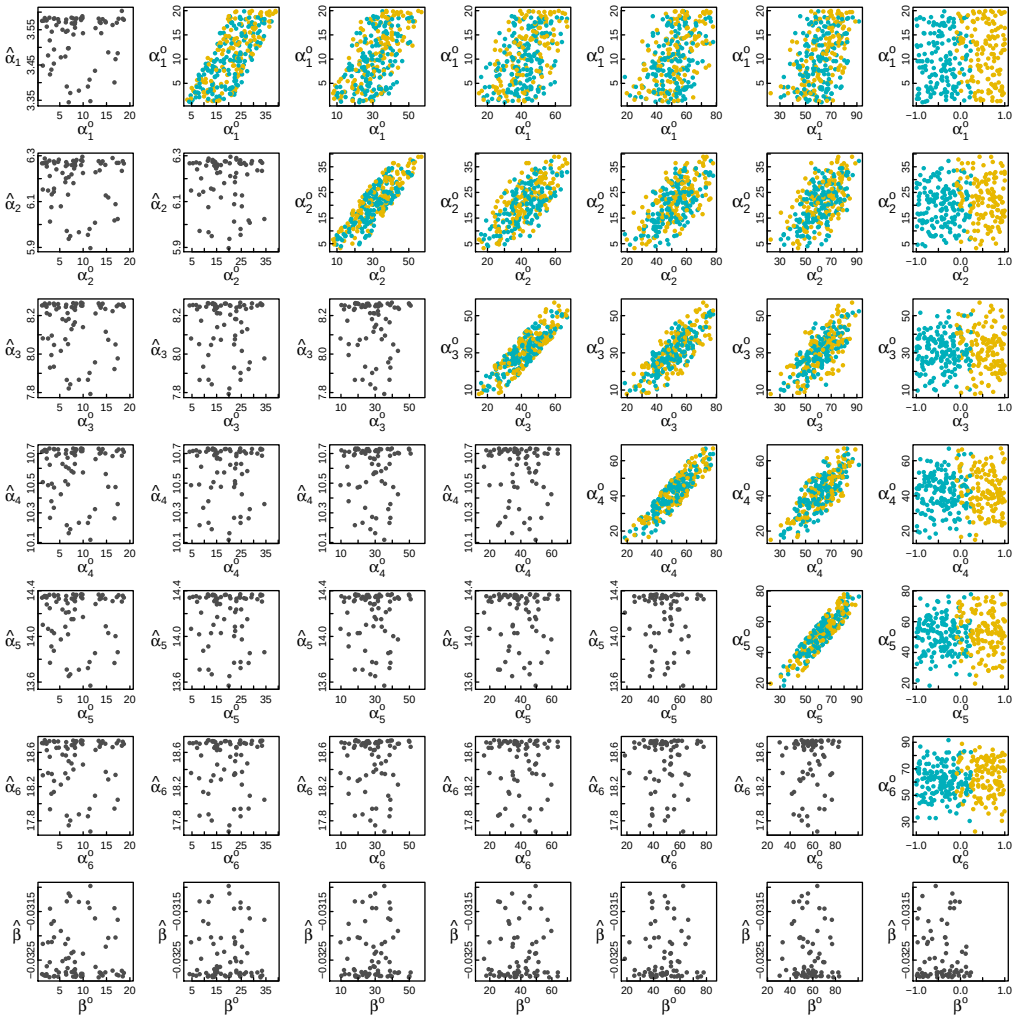

Figure S3. Sensitivity of optimisation results to initial values for the cloglog-link cumulative model with constant variance. The lower triangle and diagonal elements of the scatterplot matrix (grey symbols) show pairwise plots of filtered parameter estimates $\hat{\alpha}_i, \hat{\beta}$ against corresponding initial values $\alpha_i^\circ, \beta^\circ$. The upper triangle shows pairwise plots of the initial values for the same model parameters against each other. Green dots indicate succesfull convergence of the optimisation, yellow dots indicate convergence failures. Optimisation failed for most $\beta^\circ > 0$.