Context-dependent consequences of lagged effects in demographic models

Eric R. Scott[1], María Uriarte[1], Emilio M. Bruna[1]

[1]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida 32611-0430 USA

[2]Department of Ecology, Evolution and Environmental Biology, Columbia University 1200 Amsterdam Avenue, New York, New York 10027 USA

[3]Center for Latin American Studies, University of Florida, Gainesville, Florida 32611-5530 USA

[4]Biological Dynamics of Forest Fragments Project, INPA-PDBFF, CP 478, Manaus, Amazonas 69011-970 Brazil

(draft: 3 September 2024)

Abstract

Text of 150 words max summarizing this amazing paper.

Introduction

Paragraph 1: MM and IPM broadly used, but technical challenges remain

1. Two types of structured population models - matrix models (Leslie 1945; Caswell 2001) and integral projection models (Ellner and Rees 2006) - are fundamental frameworks used to study demography and population dynamics.
2. Their flexibility, in concert with a rapidly growing suite of software, data, and other resources (Salguero-Gómez et al. 2015; Ellner et al. 2016; Levin et al. 2021), have facilitated their use to study a wide range of topics in ecology, evolution, and conservation (Morris and Doak 2002; Crone et al. 2011; Ellner et al. 2016).
3. While recent mathematical and statistical developments have expanded the scope of these applications (Rees and Ellner 2016; e.g., Brooks et al. 2019), a number of technical challenges remain.
4. While many of these relate to general issues surrounding model development and validation (e.g., Williams et al. 2012), others stem from the fundamental challenge of using mathematical models - albeit highly advanced ones - to describe and analyze complex biological systems.

Paragraph 2: Challenge of lagged effects

1. Once such biological challenge is the implementation of alagged effects. It has long been recognized that there is the potential for lagged effects.
2. Including lagged effects in models has been a major technical challenge, but there are now multiple approaches for doing so.
3. The studies assessing the potential for lagged effects on vital rates find that they indeed appear to be prevalent (Evers et al. 2021; Scott et al. 2022).

Paragraph 3: Why does it matter when modeling?

1. Lagged effects could be different in different habitat types.
2. This could be a big reason for differences between habitats.
3. Could also be important for stochastic vs. deterministic models.
4. Here we…
5. We conducted these analyses with both deterministic and stochastic IPMs.

Methods

Demographic methods and data

Overview of the heliconia project and data

Construction of Integral Projection Models

In preliminary investigation we found that the survival and growth of plants was better explained by treating seedlings and mature plants separately. Seedlings are physiologically different from small plants because they necessarily lack the underground reserves (of carbohydrates and meristems) that a small, mature plant may have. Therefore, we used general IPMs to model population dynamics with seedlings treated as a separate discreet class not structured by size. General IPMs allow for combinations of continuous and discrete states and transitions between them (Ellner et al. 2016).

We built three classes of IPMs for comparison which each required different functional forms of their underlying vital rates models. The simplest IPM was a general, density-independent, deterministic IPM with four sub-kernels: growth and survival (P, Equation 4), fecundity (F, i.e. production of new seedlings, Equation 5), probability of staying a seedling (always 0), and recruitment (R, i.e. seedling survival and establishment, Equation 3) (Equation 1, Equation 2, Figure 1). The probability of staying a seedling, was always equal to zero, since our definition of seedlings was first year plants only.

$$n(z', t+1) = R(z')n_s(t) + \int_L^U P(z', z)n(z, t)\,dz \tag{1}$$

$$n_s(t+1) = \int_L^U F(z)n(z, t)\,dz \tag{2}$$

$$R(z') = s_s G_s(z') \tag{3}$$

$$P(z', z) = s(z)G(z', z) \tag{4}$$

$$F(z) = p_f(z)f(z)g \tag{5}$$

The number and size of mature plants in the next census is determined by seedlings entering the mature plant population (i.e. recruitment) and survival and growth (or regression) of mature plants (Equation 1). Seedlings survive (s_s) and grow into mature plants of a particular size ($G_s(z')$) (Equation 3). Mature plants survive as a function of size ($s(z)$), and grow (or regress) to a new size as a function of their previous size ($G(z', z)$) (Equation 4). Mature plants flower with a probability that is a function of size ($p_f(z)$) and produce a number of seeds as a function of size ($f(z)$), which germinate and establish as seedlings with probability g (Equation 5).

Vital rate models for growth ($G_s(z')$ and $G(z', z)$) , survival (s_s and $s(z)$), and flowering ($p_f(z)$) were fit using the long term demographic dataset. For established plants, these three vital rates were modeled as a smooth function of size in the previous census using generalized additive models (GAMs) fit with the `mgcv` package(Wood 2011) in R version 4.4.1 (2024-06-14) (R Core Team 2020) . For consistency, seedling survival and growth were also modeled using GAMs, but without size in the previous census as a predictor (i.e. intercept only models). For growth models ($G_s(z')$ and $G(z', z)$) a scaled t family distribution provided a better fit to the data than a gaussian fit as the residuals were leptokurtic with a simple gaussian model.

To estimate reproduction we drew on additional data sources to estimate the number of fruits per flowering plant as a function of plant size and the number of seeds per fruit (together $f(z)$). Germination and establishment rates in continuous forest and forest fragments were estimated using data from ….

To build the general, density-independent, stochastic, kernel-resampled IPMs, we included environmental stochasticity in all vital rate models built using the long term demographic datset by adding a random effect of year (Figure 1). The random effect of year was included using a factor–smooth interaction which allowed the relationship between plant size and vital rates to vary in functional

form among transition years. The kernel-resampling approach is to generate kernels corresponding to each transition year in the demographic dataset using the random smooths for year, and to iterate the IPM by drawing from these randomly. This is equivalent to the matrix selection approach for matrix population models described by Caswell (2001).

For the third method, we modeled the impacts of drought on vital rates explicitly and created general, density-independent, stochastic, parameter-resampled IPMs (sensu Metcalf et al. (2015)). We calculated the standardized precipitation evapotranspiraton index (SPEI) for our site using a published gridded dataset based on ground measurements (Xavier et al. 2016) as described in Scott et al. (2022). For all vital rate models fit using the long term demographic dataset, we modeled delayed effects of SPEI using distributed lag non-linear models with a maximum lag of 36 months (Scott et al. 2022) (Figure 1). To iterate these parameter-resampled IPMs, a random sequence of SPEI values was created by sampling years of the observed monthly SPEI data. Then, 36 month lags are calculated for each year starting in February (the month of the demographic census). These values are then used to predict fitted values from the vital rates models, generating different kernels at each iteration of the IPM. With this method, the kernels of successive iterations are not entirely independent because the SPEI values used in calculating vital rates include values used in the previous two iterations, but they are ergodic.

All IPMs were constructed and iterated using the `ipmr` package in R (Levin et al. 2021). The IPMs used 100 meshpoints and the midpoint rule for calculating kernels . For each type of IPM we iterated the model for 1000 time steps, discarding the first 100 time steps to omit transient effects. Stochastic growth rates (λ_s) were calculated as the average $ln(\lambda)$ from each time step (Caswell 2001) and back-transformed to be on the same scale as deterministic lambdas for comparison. We used the distribution of established plant sizes and proportion of seedlings from the full dataset as a starting population vector. While other starting population vectors were possible, the choice is of little importance as it will only impact transient dynamics, which we aren't interested in for this study.

To estimate uncertainty around the per-capita growth rates (lambdas), we created 500 bootstraps of the demographic dataset by sampling individual plants with replacement within each habitat. For each bootstrap, we then re-fit vital rates models (all except germination and establishment rate, fruits per flowering plant, and seeds per fruit, which were estimated using different datasets), constructed IPMs, and calculated a value for lambda as described above. We then used these bootstraped estimates of lambda to calculate bias corrected 95% confidence intervals (Ellner et al. 2016).

This workflow was managed using the `targets` R package (Landau 2021) which also allowed us to track computational time spent on each IPM for comparison.

Statistical analyses

All about the stats.

Results & Discussion

For all vital rates estimated using the long term demographic dataset, the DLNM model fit the best (dAIC = 0) followed by the model with a random effect of year, followed by the deterministic model (Table 1).

Population growth rates were consistently higher in continuous forest compared to forest fragments across IPM types (Table 3).

We found that the the choice of Integral Projection Model didn't change the relative ranking of lambda in Continous Forest and Fragments.

The time to iterate the DLNM models is much higher than than deterministic and kernel-resampled.

The greater use of computational resources is likely a result of `predict()` being much slower for GAMs with 2D smooths because the number of knots is much higher compared to the GAMs used for the vital rates models in the determinsitic and kernel-resampled IPMs.

Figure 2 has some interesting things in it:

- for the deterministic IPM (and the kernel-resampled IPM?) there are slightly more of the smallest plants and the largest plants in CF compared to FF (i.e. more medium sized plants in FF).

- For the kernel-resampled IPM (random effect of year), the fluctuations are extremely similar between CF and FF

- For the parameter-resampled IPM (DLNM) the size structure of the population is a LOT more variable in FF. This makes sense as we know lagged effects are more important in fragments.

- Also, the fluctuations in size structure in CF do not match the fluctuations in FF as well (can see this by the increased spread of points in Figure 2 (B)

- Also, in the parameter-resampled IPM (and only in this one), we see a shift toward smaller plants in FF compared to CF

These results are consistent with those of Kaye and Pyke (2003), who found who found that the method effected stochastic lambda but relative ranking of populations was consistent.

CRediT Statement

ERS contributed to the conceptualization, methodology, formal analysis, and led the writing of the original draft. EMB contributed to the conceptualization, methodology, and writing and also acquired funding.

Data Availability Statement

Data and R code used in this study are archived with Zenodo at .

Literature Cited

Brooks, M. E., K. Kristensen, M. R. Darrigo, P. Rubim, M. Uriarte, E. Bruna, and B. M. Bolker. 2019. Statistical modeling of patterns in annual reproductive rates. Ecology 100.

Caswell, H. 2001. Matrix population models: Construction, analysis, and interpretation. Sinauer Associates, Sunderland.

Crone, E. E., E. S. Menges, M. M. Ellis, T. Bell, P. Bierzychudek, J. Ehrlen, T. N. Kaye, et al. 2011. How do plant ecologists use matrix population models? Ecology Letters 14:1–8.

Ellner, S. P., D. Z. Childs, and M. Rees. 2016. Data-driven modelling of structured populations: A practical guide to the integral projection model. Springer Science+Business Media, New York, NY.

Ellner, S. P., and M. Rees. 2006. Integral projection models for species with complex demography. American Naturalist 167:410–428.

Evers, S. M., T. M. Knight, D. W. Inouye, T. E. X. Miller, R. Salguero-Gómez, A. M. Iler, and A. Compagnoni. 2021. Lagged and dormant season climate better predict plant vital rates than climate during the growing season. Global Change Biology 27:1927–1941.

Kaye, T. N., and D. A. Pyke. 2003. The effect of stochastic technique on estimates of population viability from transition matrix models. Ecology 84:1464–1476.

Landau, W. M. 2021. The targets R package: A dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing 6:2959.

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. Biometrika 33:183–212.

Levin, S. C., D. Z. Childs, A. Compagnoni, S. Evers, T. M. Knight, and R. Salguero-Gómez. 2021. Ipmr: Flexible implementation of Integral Projection Models in R. Methods in Ecology and Evolution 12:1826–1834.

Metcalf, C. J. E., S. P. Ellner, D. Z. Childs, R. Salguero-Gómez, C. Merow, S. M. McMahon, E. Jongejans, et al. 2015. Statistical modelling of annual variation for inference on stochastic population dynamics using Integral Projection Models. Methods in Ecology and Evolution 6:1007–1017.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology: Theory and practice of population viability analysis. Sinauer, Sunderland, MA.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria.

Rees, M., and S. P. Ellner. 2016. Evolving integral projection models: Evolutionary demography meets eco-evolutionary dynamics. Methods in Ecology and Evolution 7:157–170.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, Y. M. Buckley, J. Che-Castaldo, H. Caswell, D. Hodgson, et al. 2015. The COMPADRE Plant Matrix Database: An open online repository for plant demography. (M. Rees, ed.)Journal of Ecology 103:202–218.

Scott, E. R., M. Uriarte, and E. M. Bruna. 2022. Delayed effects of climate on vital rates lead to demographic divergence in Amazonian forest fragments. Global Change Biology 28:463–479.

Williams, J. L., T. E. Miller, and S. P. Ellner. 2012. Avoiding unintentional eviction from integral projection models. Ecology 93:2008–2014.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models 73:3–36.

Xavier, A. C., C. W. King, and B. R. Scanlon. 2016. Daily gridded meteorological variables in Brazil (1980–2013). International Journal of Climatology 36:2644–2659.

Table 1: Comparison of vital rate models used to build IPM. The 'Effect of Environment' column describes how environmental effects were included in models. Those with 'none' were used to build deterministic IPMs; those with a random effect of year were used to build stochastic, kernel-resampled IPMs; and those with a distributed lag non-linear model (DLNM) were used to build stochastic, parameter-resampled IPMs. 'edf' is the estimated degrees of freedom of the penalized GAM. ΔAIC is calculated within each habitat and vital rate combination. ΔAIC within 2 indicates models are equivalent.

Habitat	Vital Rate	Effect of Environment	edf	ΔAIC
CF	Survival	Random effect of year	43.26	0
CF	Survival	DLNM	19.72	78.92
CF	Survival	None	4.976	260
CF	Growth	Random effect of year	78.43	0
CF	Growth	DLNM	23.87	158.5
CF	Growth	None	7.81	1896
CF	Flowering	DLNM	19.59	0
CF	Flowering	Random effect of year	17.19	1.627
CF	Flowering	None	7.468	381.9
CF	Seedling survival	None	1	0
CF	Seedling survival	Random effect of year	1.817	1.386
CF	Seedling survival	DLNM	4.008	1.528
CF	Seedling growth	Random effect of year	9.475	0
CF	Seedling growth	DLNM	8.952	2.902
CF	Seedling growth	None	1	172.3
FF	Survival	DLNM	14.95	0
FF	Survival	Random effect of year	19.21	35.68
FF	Survival	None	4.333	51.25
FF	Growth	DLNM	25.18	0
FF	Growth	Random effect of year	37.84	200
FF	Growth	None	5.599	382.8
FF	Flowering	DLNM	20.61	0
FF	Flowering	Random effect of year	13.81	27.4
FF	Flowering	None	5.007	101.7
FF	Seedling survival	DLNM	5.574	0
FF	Seedling survival	Random effect of year	5.088	5.721
FF	Seedling survival	None	1	6.491
FF	Seedling growth	Random effect of year	6.25	0
FF	Seedling growth	DLNM	8.182	2.29
FF	Seedling growth	None	1	5.745

Table 2: Figure caption to be written

IPM Type	mean time (min.)
Deterministic	0.02
Stochastic, kernel-resampled	0.07
Stochastic, parameter-resampled	87.12

Table 3: Population growth rates for continuous forest (CF) and forest fragments (FF) under different kinds of IPMs with bootstrapped, bias-corrected, 95% confidence intervals.

IPM	Habitat	λ
Deterministic	FF	0.9778 (0.9736, 0.9823)
Deterministic	CF	0.9897 (0.9877, 0.9920)
Stochastic, kernel-resampled	FF	0.9787 (0.9735, 0.9835)
Stochastic, kernel-resampled	CF	0.9913 (0.9892, 0.9939)
dlnm	FF	0.9595 (0.9459, 0.9689)
dlnm	CF	0.9795 (0.9752, 0.9867)
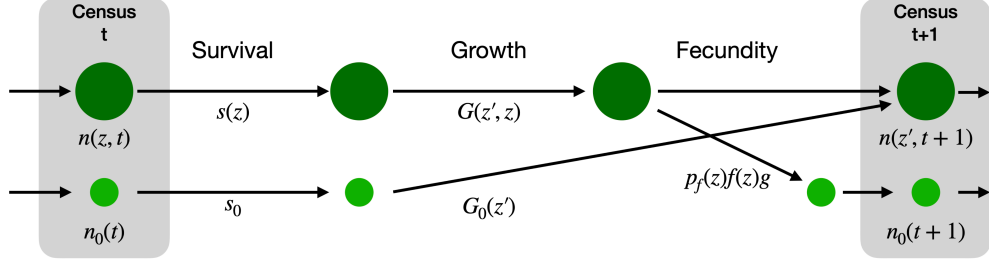
Description	Deterministic	Stochastic, kernel-resampled	Stochastic, parameter-resampled
Survival	$s(z)$	$s_y(z)$	$s(z; \theta_{0-36})$
Growth	$G(z'; z)$	$G_y(z'; z)$	$G(z', z; \theta_{0-36})$
Flowering	$p_f(z)$	$p_{f_y}(z)$	$p_f(z; \theta_{0-36})$
Size-specific fecundity	$f(z)$	$f(z)$	$f(z)$
Germination & establishment	g	g	g
Seedling survival	s_0	s_{0_y}	$s_0(\theta_{0-36})$
Seedling growth	$G_0(z')$	$G_{0_y}(z')$	$G_0(z'; \theta_{0-36})$

Figure 1: Lifecycle diagram of Heliconia acuminata. Each transition is associated with an equation for a vital rate function. The functions shown on the diagram correspond to those used to construct a general, density-independent, deterministic IPM. The table below shows the equivalent equations for stochastic, kernel-resampled IPMs and stochastic, parameter-resampled IPMs.
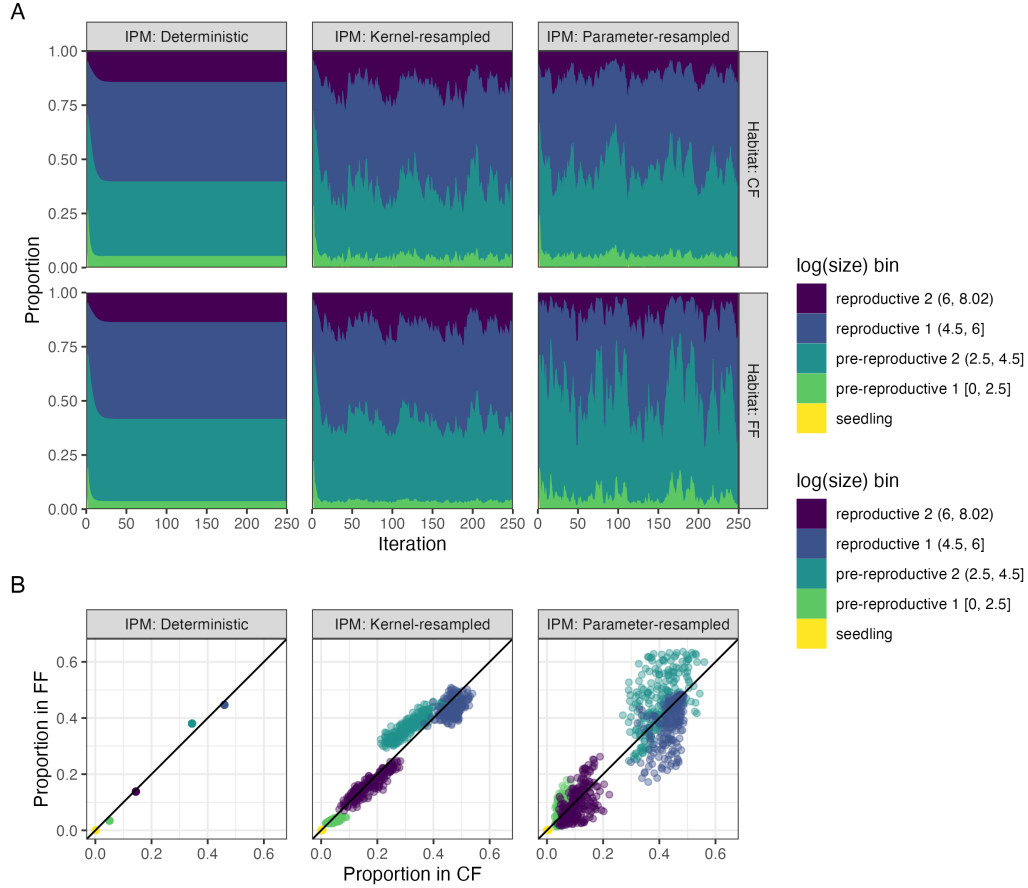
" "

Figure 2: Relative proportions of plant sizes in the first 250 iterations of the IPM simulations. Stacked area charts (A) show the relative size/stage distribution of plants in continuous forest (CF, top row) and forest fragments (FF, bottom row) in each of the three IPMs (columns). The proportion of each size class in CF and FF for each iteration is shown in B with the first 30 iterations removed to not include transient dynamics. A 1:1 line is plotted in black. Size categories include seedlings (a discrete category in the IPMs), pre-reproductive 1 (log(size) 0–2.5) that have low average survival (< 0.9) and a near 0 probability of flowering, pre-reproductive 2 (log(size) 2.5–4.5) that have a higher average survival probabilty (> 0.8) and a near 0 probability of flowering, reproductive 1 (log(size) 4.5–6) that have a high average survival probability (>0.95) and a lower flowering probability (< 0.25), and reproductive 2 (log(size) 6+) that have a high average survival probability (>0.95) and higher flowering proability (> 0.2).