

2012

# Estimating covariation between vital rates: a simulation study of connected vs. separate generalized linear mixed models (GLMMs)

Margaret EK Evans

Kent E. Holsinger

*University of Connecticut - Storrs*, [kent.holsinger@uconn.edu](mailto:kent.holsinger@uconn.edu)

Follow this and additional works at: [https://opencommons.uconn.edu/eeb\\_articles](https://opencommons.uconn.edu/eeb_articles)

 Part of the [Population Biology Commons](#)

## Recommended Citation

Evans, Margaret EK and Holsinger, Kent E., "Estimating covariation between vital rates: a simulation study of connected vs. separate generalized linear mixed models (GLMMs)" (2012). *EEB Articles*. 30.  
[https://opencommons.uconn.edu/eeb\\_articles/30](https://opencommons.uconn.edu/eeb_articles/30)

1 Invited contribution to a special issue of Theoretical Population Biology:

2 “Developments in structured models: construction, analysis, and inference”

3

4 **Estimating covariation between vital rates: a simulation study of connected vs. separate**  
5 **generalized linear mixed models (GLMMs)**

6

7 **Abstract**

8 Covariation between vital rates is recognized as an important pattern to be accounted for in  
9 demographic modeling. We recently introduced a model for estimating vital rates and their  
10 covariation as a function of known and unknown effects, using generalized linear mixed  
11 models (GLMM's) implemented in a hierarchical Bayesian framework (Evans et al. 2010). In  
12 particular, this model included a model-wide year effect (YEAR) influencing all vital rates,  
13 which we used to estimate covariation between vital rates due to exogenous factors not  
14 directly included in the model. This YEAR effect connected the GLMMs of vital rates into  
15 one large model; we refer to this as the “connected GLMMs” approach. Here we used a  
16 simulation study to evaluate the performance of a simplified version of this model,  
17 compared to separate GLMMs of vital rates, in terms of their ability to estimate correlations  
18 between vital rates. We simulated data from known relationships between vital rates and a  
19 covariate, inducing correlations among the vital rates. We then estimated those  
20 correlations from the simulated data using connected vs. separate GLMMs with year  
21 random effects. We compared precision and accuracy of estimated vital rates and their  
22 correlations under three scenarios of the pervasiveness of the exogenous effect (and thus  
23 true correlations). The two approaches provide equally good point estimates of vital rate  
24 parameters, but connected GLMMs provide better estimates of covariation between vital

25 rates than separate GLMMs, both in terms of accuracy and precision, when the common  
26 influence on vital rates is pervasive. We discuss the situations where connected GLMMs  
27 might be best used, as well as further areas of investigation for this approach.

28

29 Margaret E. K. Evans<sup>1,3</sup>, Kent E. Holsinger<sup>2</sup>

30

31 1. Origin, Structure and Evolution of Biodiversity UMR 7205, Museum National d'Histoire  
32 Naturelle, 16 rue Buffon, 75231 Paris cedex 05

33 2. Department of Ecology & Evolutionary Biology, University of Connecticut, U-3043, Storrs,  
34 CT 06269-3043

35 3. author for correspondence: Margaret.ekevans@gmail.com

36

37

38

39 Keywords: vital rate covariation, generalized linear mixed models, random effects, year  
40 effects, hierarchical Bayesian model, demography, transition matrix model, estimation

41

42

43

44

45

## 46 **1. Introduction**

47 Covariation among vital rates is recognized as an important pattern to be accounted  
48 for in demographic modeling (van Tienderen 1995, Morris and Doak 2002, Boyce et al.  
49 2006). For those vital rates with a positive effect on population growth rate, positive  
50 covariation between vital rates amplifies the effects of environmental variation, reducing  
51 population growth rate, whereas negative covariation has the opposite effect (Tuljapurkar  
52 1990, Tuljapurkar 2009, Tomimatsu and Ohara 2010). A failure to account for covariation  
53 between vital rates can yield misleading results about what part of the life cycle should be  
54 targeted for conservation (i.e., sensitivity analysis; Sim et al. 2011) as well as what the risk of  
55 population extinction is (Doak et al. 1994, Morris and Doak 2002). Other consequences  
56 include the fact that increased variation in a given vital rate doesn't always reduce the  
57 population growth rate: an increase in the variation of less influential vital rates can lead to  
58 an increase in the population growth rate, if those rates covary negatively with more  
59 influential vital rates (Orzack and Tuljapurkar 1989, Doak et al. 2005, Haridas and  
60 Tuljapurkar 2005).

61 Evidence from natural populations also suggests that covariation between vital rates  
62 can have substantial effects. Coulson et al. (2005) found that covariation between  
63 demographic rates was responsible for about one third of all variation in population growth  
64 in three ungulate populations. Similarly, Ezard et al. (2006) found that vital rate covariation  
65 explained 25-30% of the variation in growth rate of a population of the common tern. In  
66 contrast, Morris et al. (2011) found only weak effects of vital rate covariation on variation in  
67 population growth in seven primate species, but they argue this is an exceptional pattern  
68 (compared to 23 other vertebrates). Jongejans et al. (2010) found by analyzing data from 40  
69 plant species that positive covariation between reproduction and survival predominated

70 (rather than negative covariation) and concluded that increased climate variation caused by  
71 anthropogenic climate change could result in increased variation in plant population  
72 dynamics and increased extinction risk (via this effect of positive covariation).

73         With these important effects in mind, we recently introduced a model for estimating  
74 vital rates and their covariation as a function of known and unknown effects, using  
75 generalized linear mixed models (GLMM's; Evans et al. 2010). Our model was inspired by  
76 the endangered plant that we studied (*Dicerandra frutescens*): time-since-fire is known to  
77 influence demography throughout its life cycle, inducing correlations among vital rates. At  
78 the same time, we supposed that weather variation may cause a second layer of variation  
79 and covariation of vital rates. Analogous to a random block effect in an experimental  
80 setting, we modeled these unmeasured environmental effects with a random year effect.  
81 What makes this year effect (YEAR) different is the fact that it is shared by all of the  
82 generalized linear mixed models of vital rates. Thus the model analyzed all parts of the life  
83 cycle as a function of time-since-fire and year variation simultaneously (with additional  
84 parameters estimating the sign and magnitude of the YEAR effect on a given vital rate  
85 compared to a baseline vital rate). This allowed us to estimate (and incorporate into  
86 subsequent simulations of population dynamics) positive or negative covariation among  
87 vital rates arising both through systematic effects of an exogenous variable and through  
88 random effects associated with particular years. We refer to this as the "connected  
89 GLMMs" approach to estimating covariation between vital rates.

90         Here we compare the "connected GLMMs" approach to a simpler model structure,  
91 where each vital rate is estimated via a separate GLMM, each with a random year effect  
92 (which we refer to as the "separate GLMMs" approach). We know of no example in the  
93 literature where this exact structure has been adopted; however, over last ten years, vital

94 rates have increasingly come to be modeled with generalized linear mixed models (*e.g.*,  
95 Gross et al. 2006, Le Galliard et al. 2010). This marks a change from the methods (for  
96 estimating vital rates, and simulating their variation and covariation) recommended in  
97 authoritative texts on the subject ten years ago (Caswell 2001; Morris and Doak 2002).

98         We compare the ability of the connected vs. separate GLMM approaches to estimate  
99 vital rates and their correlations using simulated data. Specifically, we ask if one approach  
100 performs better than the other, in terms of either accuracy or precision, at estimating vital  
101 rates and correlations among vital rates. Or, are there different conditions under which  
102 each approach performs better? *A priori*, we might expect the “connected GLMMs”  
103 approach to estimating vital rates and their covariation to perform best when many parts of  
104 the life cycle experience strong common influence. Conversely, we might expect the  
105 “separate” GLMMs approach to perform as well as, or perhaps even better than the  
106 “connected” approach, when there are few vital rates experiencing common influence, or  
107 that influence is weak. To address these questions, we simulated data from known  
108 relationships between vital rates and a covariate (time-since-fire), inducing correlations  
109 among the vital rates. As a check of our workflow, we first estimated vital rates and their  
110 correlations using a model that reflects the process generating the simulated data. We then  
111 estimated vital rate correlations from the simulated data using models that treat the  
112 covariate as unknown. This was intended to mimic a realistic scenario of estimation: an  
113 unknown covariate affects vital rates and is modeled with a random year effect. We  
114 considered two aspects of model performance under three profiles with respect to the  
115 strength of covariation and the parts of the life cycle affected: precision and accuracy of the  
116 (1) estimated vital rates and (2) correlations between vital rates.

117

118 **2. Methods**

119 *Data simulation* – We simulated data under a simplified version of the life history and  
120 ecology of *Dicerandra frutescens*, a fire-dependent, endangered plant endemic to Florida  
121 scrub that has been modeled with a six by six transition matrix (Menges et al. 1999, 2006,  
122 Evans et al. 2010; see Appendix 1). We used the posterior mean estimates of vital rates and  
123 time-since-fire effects inferred in Evans et al. (2010) as the “true” demography of our  
124 simulated species (Profile 1). Fire kills plants but not seeds, the latter remaining viable in  
125 the soil more than one year, thus populations recover from fire via germination from a seed  
126 bank. In general, time-since-fire negatively affects the demography of this simulated  
127 version of *D. frutescens*: germination declines, seed production declines, transitions forward  
128 (in terms of size or developmental stage) decline, and transitions backwards increase  
129 (Profile 1). Exceptions include a positive effect of time-since-fire on initial seedling survival  
130 (survival from the time of seedling emergence until the time of the annual census; Profile 1).

131 Data simulation included four stochastic components in the life cycle (seed fates,  
132 initial seedling fate, plant fates, and seed production) in addition to stochastic fire history.  
133 Each simulation started with the year of fire (tsf=0) and a population consisting of  $10^5$  seeds.  
134 At each subsequent time step, the occurrence of fire was a coin flip, with increasing  
135 probability as time-since-fire increased, according to the cumulative distribution function of  
136 a Weibull distribution (with median interval between fires of 14 years and a shape  
137 parameter of 32). Vital rates were calculated as a function of time-since-fire (using the  
138 equations detailed below and in Appendix A), and the multinomial or binomial probabilities  
139 governing seed, seedling, and plant fates were calculated from these vital rates. Seed fates  
140 were stochastic draws from a multinomial distribution with three outcomes: a seed (1)  
141 survives and doesn't germinate; a seed (2) survives and germinates; or a seed (3) neither

142 germinates nor survives. Initial seedling fate was drawn from a binomial model of survival  
 143 from the time of germination to the time of the annual census. The fates of five classes of  
 144 plants (seedling, non-flowering plant, and three size-based classes of flowering plants: small,  
 145 medium, and large) were stochastic draws from multinomial distributions with five  
 146 outcomes: non-flowering plant, small, medium, or large flowering plants, or dead. For each  
 147 class of flowering plants, seed production was a stochastic draw from a Poisson distribution.  
 148 Fifteen replicate populations were simulated in this way for 20 time steps each.

149 The probabilities governing the binomial, multinomial, and Poisson distributions  
 150 depended on the natural log transform of time-since-fire (lnTSF; since vital rates of  
 151 *Dicerandra frutescens* were better predicted by linear regression of lnTSF than TSF). For  
 152 example, the logit of the probability of initial seedling survival was a function of ln-  
 153 transformed TSF:

$$154 \quad \text{logit}(s_{[t]}) = \beta_0^s + \beta_1^s \cdot \ln(TSF_{[t]})$$

155 where  $\beta_0^s$  is an intercept term (the baseline value of the vital rate when TSF=0) and  $\beta_1^s$  is the  
 156 effect of ln-transformed TSF on seedling survival. Similarly, germination fraction (g), which  
 157 enters in the multinomial governing seed fates, was a function of ln-transformed TSF:

$$158 \quad \text{logit}(g_{[t]}) = \beta_0^g + \beta_1^g \cdot \ln(TSF_{[t]})$$

159 where  $\beta_0^g$  is an intercept term and  $\beta_1^g$  is the effect of ln-transformed TSF on germination.

160  
 161 The probabilities of transitions among plant classes, which are multinomial responses,  
 162 depended on ln-transformed TSF as well. For example, the probability of a seedling  
 163 becoming a vegetative plant ( $a_{[3,2]}$ ) is proportional to



165

166 
$$\log(a_{[t,3,2]}) \propto \beta_{0[3,2]}^{tr} + \beta_1^{P1} \cdot \ln(TSF_{[t]})$$

167

168 The parameter  $\beta_{0[3,2]}^{tr}$  is the intercept and  $\beta_1^{P1}$  is the effect of ln-transformed TSF on a one-  
169 step progression forward. Note that a logit transformation, of the general form  $\log(p/(1-p))$ ,  
170 is possible only for a binomial response (e.g., initial seedling survival and seed germination,  
171 above). For a multinomial response, the analog is either a baseline logit, which we use here,  
172 or a cumulative logit. For a baseline logit, we set  $\log(\phi_i) = \beta_{0[i]}^{tr} + \beta_1 \cdot \ln(TSF)$  for  $i = 1, \dots, n-1$   
173 and  $\log(\phi_n) = 0$ , where category  $n$  is the baseline. We selected the category death as the  
174 baseline. Then  $a_i = \phi_i / \sum(\phi_i)$ . Another way of saying this is that for each class of plant  
175 (state at time  $t$ ), we standardized the odds of other fates to the odds of death ( $\log(\phi_{t,7,j}) = 0$ ;  
176 see Appendix A), and thus the probability of death is  $1 / \sum_i \phi_{t,i,j}$ . The equations governing  
177 all 25 plant transitions are detailed in Appendix A, Table 1.

178 We simulated seed production with a Poisson process dependent on time-since-fire.  
179 Because we chose to simplify our study by not estimating the parameters involved in seed  
180 production, we provide less detail here on this Poisson process (see Evans et al. 2010 for  
181 details).

182 We created a series of demographic profiles, with progressively fewer and weaker  
183 effects of time-since-fire on vital rates, against which to test the alternative statistical  
184 models (Table 1). In the first profile, all vital rates were influenced by time-since-fire, using  
185 the posterior mean estimate of time-since-fire effects from Evans et al. (2010), as described  
186 above. In the second profile, we set to zero those time-since-fire effects with 95% central  
187 density overlapping zero in Evans et al. (2010), yielding nine non-zero time-since-fire effects

188 (Table 1). In the third profile, we retained just the four strongest non-zero time-since-fire  
189 effects, and set these effects to 10% of their posterior mean values (Table 1). As fewer vital  
190 rates are affected by time-since-fire the true correlations between vital rates become  
191 correspondingly few.

192

193 << Table 1 approximately here >>

194

195 *Estimation* – We compared the ability of three statistical models to estimate vital rate  
196 parameters and correlations between vital rates. The first of these statistical models  
197 reflects exactly the above equations (and the equations in Table 1 in Appendix A), which  
198 were used to simulate the data (providing a check of our models and scripts). We refer to  
199 this as the “TSF” model (Table 2), since it includes the true time-since-fire (TSF) data and  
200 estimates the effects of ln-transformed TSF on vital rates ( $\beta_1$  parameters above).

201 The second statistical model treats time-since-fire as unknown, and instead  
202 estimates a random year effect shared among vital rates (YEAR), in addition to independent  
203 year effects unique to each vital rate. We refer to this as the “YEAR” model (Table 2). It  
204 reflects the “connected GLMMs” approach to estimating vital rates and their covariation,  
205 introduced in Evans et al. 2010. To ensure that the model parameters were well-identified,  
206 we assigned a positive coefficient of one to the YEAR effect in the regression for the  
207 probability of a large flowering plant becoming a vegetative plant. Thus, the YEAR effect can  
208 be thought of as year variation in the transition from large flowering to vegetative, and year  
209 variation in all other vital rates is estimated relative to this baseline vital rate. Associated  
210 with the YEAR effect are coefficients ( $\beta_2$  parameters, Table 2; see also Appendix A, Table 1),  
211 modifying the sign of the YEAR effect on a given vital rate relative to this baseline vital rate

212 (as described in Evans et al. 2010). Additional “independent” year effects ( $\epsilon_{yr}$ ) modify the  
213 magnitude of the YEAR effect on a given vital rate compared to the baseline vital rate. Thus  
214 the effect of year variation on each vital rate is free to vary in both sign and magnitude,  
215 compared to the baseline vital rate.

216 The third statistical model also treats time-since-fire as unknown, estimating year  
217 effects unique to each vital rate ( $\epsilon_{yr}$ ). This follows the “separate GLMMs” approach to  
218 estimating vital rates and their covariation. We refer to this model as the “iYear” model  
219 (year effects on each vital rate are estimated independently of one another; Table 2).

220

221 << Table 2 approximately here >>

222

223 We assigned prior distributions that were broad within a biologically reasonable  
224 range, as described in Evans et al. (2010). To improve convergence of the YEAR and iYear  
225 models, we used a method of reparameterization known as hierarchical centering (Gelfand  
226 1995), described in further detail in Appendix A (and in Evans et al. 2010).

227

228 *Model Implementation* – We sampled from the joint posterior distributions of the  
229 parameters of the three competing statistical models using Metropolis-Hastings Markov  
230 Chain Monte Carlo (MCMC) simulation in OpenBUGS (Version 3.1.2; Thomas et al. 2006).  
231 The MCMC simulations consisted of three independent chains, with initial values chosen  
232 randomly from the prior distributions. We evaluated convergence of the chains using the  
233 Gelman-Rubin (1992) diagnostic, which compares variation within vs. among chains. We  
234 discarded the burn-in period (identified by checking traces of the chains) and thinned  
235 samples based on examination of autocorrelation plots, to obtain more or less independent

236 samples from the joint posterior distribution of model parameters (Carlin et al. 2006,  
237 Congdon 2007).

238 For the TSF model, we ran the MCMC simulation for  $5.5 \cdot 10^4$  iterations, with the first  
239 5,000 iterations discarded as burn-in, and retained every 50<sup>th</sup> sample. The YEAR model  
240 (especially the parameters associated with the multinomial model of plant transitions) took  
241 longer to converge and required stronger thinning: we discarded the first  $10^5$  iterations as  
242 burn-in and retained every 100<sup>th</sup> sample from a second  $10^5$  iterations. The iYear model took  
243 even longer to converge, but autocorrelation was more limited: we discarded the first  
244  $1.5 \cdot 10^5$  iterations as burn-in and retained every 50<sup>th</sup> sample from another  $5 \cdot 10^4$  iterations.

245

246 *Model Evaluation* – We evaluated the three statistical models in terms of their ability to  
247 estimate parameters underlying vital rates and correlations among vital rates. Taking the  
248 example of initial seedling survival ( $s$ ), the TSF model estimates two parameters,  $\beta_0^s$  and  $\beta_1^s$ .  
249 The other two models (YEAR, iYear) also estimate  $\beta_0^s$ , but they do not estimate  $\beta_1^s$ , since  
250 time-since-fire is treated as unknown. We compared posterior mean estimates of  
251  $\beta_1^s \cdot \ln(TSF_{[t]})$  for each time step and replicate of population simulations (N=15  
252 replicates\*20 time steps=300) to its equivalent in the YEAR model and the iYear model: the  
253 estimate of initial seedling survival (for a given time step and replicate) minus the estimate  
254 of the intercept term for initial seedling survival:

255

$$256 \quad \text{logit } s_{[t]} - \beta_0^s = \beta_1^s \cdot \ln(TSF_{[t]}) \quad \text{TSF model}$$

$$257 \quad \text{logit } s_{[t]} - \beta_0^s = \beta_2^s \cdot YEAR_{[t]} + \varepsilon_{yr[t]}^s \quad \text{YEAR model}$$

258  $\text{logit } s_{[t]} - \beta_0^s = \varepsilon_{yr[t]}^s$

iYEAR model

259 .

260 To estimate correlations between vital rates, we first estimated the posterior mean  
261 vital rate for each time step and replicate of population simulations from each of the three  
262 statistical models. We calculated the correlation between each pair of vital rates by calling  
263 the cor function in R on the time series (20 time steps) of posterior mean vital rates. This  
264 was repeated for each replicate of population dynamics, yielding 15 estimates of pairwise  
265 correlations between vital rates. These correlations were calculated from the vital rates  
266 expressed on the probability scale rather than the log or logit scale. Note that correlations  
267 between vital rates are perfect on the log or logit scale (through their shared linear  
268 relationships with TSF). That is, the correlation between  $\text{logit}(X)$  and  $\text{logit}(Y)$  where  $\text{logit}(X)$   
269  $= a + bu$  and  $\text{logit}(Y) = c + du$  is exactly one for any random variable  $u$ . The correlation  
270 between  $X$  and  $Y$  is less than 1, because of the non-linear relationships involved, but it is still  
271 very high because the relationship is close to linear over much of the range. This property  
272 arises from any model that includes a deterministic relationship between an environmental  
273 covariate and vital rates via a generalized linear model.

274 We then compared posterior mean estimates of pairwise correlations from the three  
275 statistical models to the realized pairwise correlations (calculated by calling the cor function  
276 in R on the time series of realized vital rates) in terms of accuracy and precision. Note that  
277 demographic stochasticity can cause the realized correlation between two vital rates (for a  
278 given replicate of simulated population dynamics) to differ substantially from the value  
279 expected if the sample size were infinite (or very large). This is especially true for  
280 parameters very close to zero or one. For example, germination fraction is very low in the  
281 life history that we modeled. If there are few seeds, the realized germination fraction will

282 often be zero, and correlations between germination fraction and other vital rates differ  
283 substantially from their expectation under infinite sample size. For this reason, we  
284 compared the correlations between vital rates estimated by the three statistical models to  
285 the realized pairwise correlations.

286         Each replicate of population dynamics yielded a unique realized correlation between  
287 each pair of vital rates (depending on demographic stochasticity as well as fire history). For  
288 each replicate then, we found the difference between the realized correlation and the  
289 posterior mean correlation estimated from each of the three statistical models (bias); we  
290 report the average bias across replicates. We evaluated the precision of estimates of vital  
291 rate correlations in terms of mean root squared error: the square root of the mean squared  
292 difference between realized correlations and estimated correlations (N=15 replicates).

293         Note that we did not consider all the pairwise correlations between vital rates: we  
294 chose not to examine correlations between transitions within a column (*e.g.*, the correlation  
295 between the transition from seedling to vegetative and the transition from seedling to small  
296 flowering), since those transitions are constrained to sum to one. Further, we considered  
297 only a subset (55 total) of the correlations between transitions in different columns: all  
298 possible combinations involving stasis, a one-step progression forward (seedling to  
299 vegetative, vegetative to small flowering, etc.), or a one-step retrogression backward (small  
300 flowering to vegetative, medium flowering to small flowering, large flowering to medium  
301 flowering). We chose to focus on these transitions because they are the most commonly  
302 traversed demographic pathways in the life history.

303

304

305 **3. Results**

### 306 **3.1 Estimates of vital rate parameters**

307 A model that estimates correlations between vital rates well but does not estimate vital  
308 rates well would not be useful, so we first compare the ability of the three statistical models  
309 to estimate the parameters underlying vital rates (under Profile 1).

310

311 *TSF model vs. true parameter values* – By comparing the parameter values used for data  
312 simulation (“true” values) to the parameter estimates from the TSF model, which exactly  
313 reflects the process used to simulate the data, we validate the various components of our  
314 workflow. The parameters associated with transitions for which there are many  
315 observations are recovered well by the TSF model. For example, most of the intercept  
316 terms from the multinomial model of plant transitions are estimated well (Figure 1a), but  
317 parameters associated with transitions for which there are few observations are not  
318 estimated well, *e. g.*, seedling to large flowering, small flowering to large flowering, medium  
319 flowering to vegetative, and large flowering to vegetative and small flowering (arrows,  
320 Figure 1a). The same is true for the slope terms: the effects of time-since-fire on stasis of  
321 vegetative plants and small and medium flowering plants are estimated well, but the effect  
322 of TSF on stasis of the rarest class of plants (large flowering) is not estimated well. That is,  
323 the combined effects of small sample size and demographic stochasticity weaken our ability  
324 to detect the true effect of time-since-fire.

325

326 *All three models vs. true parameter values* – Not surprisingly, the TSF model yielded the best  
327 estimates of vital rate parameters, among the three statistical models. We focus on  
328 estimation of time-since-fire effects, or their equivalent in the YEAR and iYear models, since  
329 the intercept terms are parameterized in exactly the same manner in all three models.

330 Across all the multinomial transitions, TSF effects (or their equivalent) were most accurately  
331 estimated by the TSF model, but the YEAR and iYear models were not too far off (on  
332 average), and the latter two were essentially equivalent (Figure 2).

333         However, dissecting this result a little further reveals that there are substantially  
334 different ways of being wrong. The models with year effects (YEAR and iYear) are inherently  
335 more flexible, but they also are more parameter rich, leading to lower power and a poorer  
336 ability to detect TSF effects. This is illustrated in Figure 3: estimates of TSF effects (or their  
337 equivalent) are plotted against the true values, for each time step and replicate of simulated  
338 data, so that accurate estimates of parameters fall on the 1:1 line in each panel. Estimates  
339 from the TSF model (red symbols) fan away from zero for each replicate of data simulation,  
340 because the estimate of the TSF effect from a given stochastic realization of data differs  
341 from the true TSF effect by some amount, and this error is magnified with increasing TSF  
342 (Figure 3). Whereas in the YEAR and iYear models, the equivalent of the TSF effect is  
343 estimated on a year-by-year basis, so the difference between the estimated effect and the  
344 true value is independent from one year to the next (green and blue symbols, Figure 3). The  
345 TSF model has greater power though: it estimates just 2 parameters ( $\beta_0$  and  $\beta_1$ ), whereas  
346 the iYear model estimates a random error term for every transition in every year and the  
347 YEAR model adds an additional model-wide year term for every year. Thus, the prior  
348 distributions on the parameters have a larger effect in the year models, causing them to  
349 estimate weaker effects (their symbols cluster closer to zero on the x-axis in Figure 3). As a  
350 result of these strengths and weaknesses, the TSF model estimates relatively strong TSF  
351 effects more accurately (Figure 3a and b), whereas the year models estimate relatively weak  
352 effects more accurately (Figure 3b and c). This comparison of the three statistical models is  
353 based on the parameter values (time-since-fire effects) of Profile 1. As the true time-since-



354 fire effects become fewer and weaker (Profile 2 and 3), the year models do increasingly well  
355 (compared to the TSF model) at estimating vital rate parameters.

356

### 357 **3.2 Estimates of correlations**

358 Considering first Profile 1 (time-since-fire affects all vital rates; Table 1), we find that the TSF  
359 and YEAR models both estimate correlations between vital rates reasonably well, though  
360 the TSF model tends to overestimate correlations and the YEAR model tends to  
361 underestimate correlations (illustrated for a sample of three combinations of vital rates in  
362 Figure 4). The iYear model, which uses separate GLMM's to model vital rates,  
363 underestimates correlations more severely (Figure 4). The TSF model yields the most  
364 precise estimates of correlations between vital rates, followed by the YEAR model, followed  
365 by the iYear model's very poor precision (Figure 4). Estimates of correlations from the iYear  
366 model broadly overlap zero when the true correlations are quite strong (Figure 4).

367 Comparing bias and precision of the three statistical models across a sample of 79  
368 pairwise correlations shows that the YEAR and TSF models outperform the iYear model  
369 under Profile 1, but not Profile 2 and 3 (Figure 5). Under Profile 1, the TSF model and YEAR  
370 models over- and underestimate correlations, respectively, by approximately the same  
371 magnitude, whereas the iYear model strongly underestimates correlations. However, as the  
372 correlations between vital rates become few and weak (Profile 2 and 3), the gap in  
373 performance between the YEAR and iYear models closes: bias and precision of the two year  
374 models is broadly overlapping (Figure 5). Performance of the TSF model also deteriorates as  
375 time-since-fire effects are dropped: the TSF model loses precision under Profiles 2 and 3,  
376 and, like the year models, underestimates the few remaining correlations under Profile 3.

377

378

379 **4. Discussion**

380 This simulation study shows that the connected GLMMs approach (YEAR model)  
381 outperforms separate GLMMs of vital rates (iYear model) at estimating covariation between  
382 vital rates, both in terms of accuracy and precision, when the common influence on vital  
383 rates is pervasive. It is worth repeating that the “common influence” can differ in both sign  
384 and magnitude with respect to its effect on different parts of the life cycle. That is, the YEAR  
385 model is totally flexible in terms of the direction and magnitude of year variation in different  
386 vital rates (relative to a chosen baseline vital rate). This result leads to a series of new  
387 questions.

388

389 *What are real patterns of covariation in the wild like?* – If patterns of covariation between  
390 vital rates in wild populations are like Profile 1 (pervasive), this argues for using the  
391 connected GLMMs approach to estimating covariation, when exogenous influences are  
392 unknown. Some populations probably do experience such covariation, though in many of  
393 these cases the exogenous influence is known (e.g., fire, floods, hurricanes). We suggest the  
394 connected GLMMs approach might best be used in situations where a pervasive influence of  
395 weather is suspected, but it is not yet known exactly what dimension of weather is  
396 responsible. A number of papers over the last decade have identified either climatic drivers  
397 of demography or complex interactions between climate and other factors (Coulson et al.  
398 2001, Stenseth et al. 2002, and many papers citing these). Ecologists are increasingly  
399 interested in understanding how weather variation affects population or range dynamics, as  
400 the reality of anthropogenic climate change unfolds. A model-wide YEAR effect can stand in  
401 for a more explicit understanding of the influence of weather on demography, as a first

402 analysis. This model can also help in the process of identifying influential dimensions of  
403 climate: candidate climate variables can be regressed on the model-wide YEAR effect.

404 Other populations are likely to follow a profile more like Profile 2 or 3, with some life  
405 stages affected by exogenous variables and others free from influence. In particular, there  
406 is quite a famous literature on life stages, like seed banks or extended adult lifespan, that  
407 buffer environmental variation (Tuljapurkar 1982, Orzack and Tuljapurkar 1989, Tuljapurkar  
408 and Wiener 2000, Clauss and Venable 2000, Evans and Dennehy 2005, Venable 2007).

409 Populations with vital rate variation limited to a few life stages can be modeled with  
410 separate GLMMs as effectively as connected GLMMs. However, given that the connected  
411 GLMM's approach *did not perform worse* than the separate GLMMs approach under any of  
412 the (albeit limited) conditions that we tested, it could be argued that the connected GLMMs  
413 approach is the safe alternative.

414

415 *Model complexity* – While both the YEAR and iYear models (connected vs. separate GLMMs)  
416 are parameter-rich (compared to knowing and modeling the exogenous effect explicitly),  
417 the YEAR model estimates more parameters. Thus its weakness may be how well it  
418 performs under conditions of limited data. Here we simulated a situation with 20 years of  
419 data, quite a long time series, but only one population. Sampling from multiple sites that  
420 behave differently, it's possible to obtain approximately similarly informative data in 4 or 5  
421 years. But we simply don't know at this time whether the connected GLMMs approach can  
422 be supported with less data, say 4 or 5 years of data from a single site, and how its  
423 performance compares to the separate GLMMs approach under conditions of limited data.

424

425 *Estimation vs. projection* – Here we compared the performance of connected vs. separate  
426 GLMMs with respect to parameter estimation (vital rates and their covariation). Another  
427 area to explore is the advantage of using connected vs. separate GLMMs with respect to the  
428 accuracy and precision of population projections. We suspect that seemingly small  
429 improvements in the estimation of covariation between vital rates might play out as larger  
430 improvements with respect to population projection. Furthermore, the dramatic  
431 divergence of estimates from the TSF model as time since fire increases (Figure 3) suggests  
432 that even when an exogenous variable is well understood, a connected GLMM could  
433 produce population projections that are more accurate. We suggest that a model with a  
434 random YEAR effect is worth exploring (in terms of fit to the data) even when an exogenous  
435 variable is known to affect populations.

436

437 *How important is the magnitude of correlations?* – As noted above (Methods), correlations  
438 between vital rates are expected to be strong (though not perfect) when those vital rates  
439 vary solely according to a common influence. The magnitude of the exogenous effects  
440 (here,  $\beta_1$  parameters) has some effect on the strength of correlations between vital rates,  
441 but this influence is likely to be small, since the relationship between the exogenous  
442 variable and transition rates is close to linear over much of the range of the exogenous  
443 variable. Our experimental design did not allow us to explore the influence of the  
444 magnitude of the exogenous effects, since the difference between Profile 2 and 3 includes  
445 both changes to the number of vital rates affected by time-since-fire and the magnitude of  
446 those effects. Future work could examine how the magnitude of correlations between vital  
447 rates influences the performance of connected vs. separate GLMMs, either by varying the

448 magnitude of exogenous effects in isolation, or by introducing another, uncorrelated source  
449 of variation (noise) into the data simulation stage.

450

451 *Implementation* – We can offer a few words of advice (and encouragement) about  
452 implementation of the connected GLMMs model. We tied the model-wide YEAR effect to  
453 the transition from a large flowering plant to a vegetative plant (such that the YEAR effect  
454 estimated variation among years in the magnitude of this transition). Retrospectively, we  
455 realize this was not the best possible choice. Instead, we would recommend tying the YEAR  
456 effect to a vital rate for which there is a large number of observations (and a vital rate which  
457 seems to be variable from year to year). The parameters involved in the transition from  
458 large flowering to vegetative were not estimated well (see Figure 1 for the intercept term;  
459 time-since-fire effect not shown), probably because of the small number of individuals in  
460 the large flowering class as time-since-fire increases. In spite of this, the YEAR model did  
461 remarkably well at estimating known parameter values (vital rates and their covariation),  
462 which is quite encouraging.

463

464 *Generality* – We should point out that the results obtained here do not depend on the  
465 method of parameter estimation, the specific life history structure, or the specific  
466 environmental factor. Similar results should be obtained whether likelihood or Bayesian  
467 methods are used for parameter estimation. Similar results should also be obtained with a  
468 different life history. That is, the exact values of vital rates and the dimensionality of the  
469 projection matrix might be different in another life history, but this has no effect on the  
470 rules of how correlations between vital rates arise (given a common influence on vital rates  
471 modeled in a GLMM framework). While fire as a form of disturbance has the special feature

472 of creating a sequence or cycle of changing conditions (rather than randomly changing  
473 conditions), this property does not enter into the GLMMs used to estimate vital rates, thus  
474 TSF could just as easily be any number of other continuous variables with a monotonic  
475 relationship with vital rates (e.g., total solar flux or total soil fertility).

476         Looking forward, one of the important tools emerging in population modeling is  
477 integral projection models, where population structure variables are continuous (or a  
478 mixture of continuous and categorical; Easterling et al. 2000). IPMs use GLMMs to estimate  
479 vital rates, so, while we're not aware of any case where this has been done, random year  
480 effects can be added to the GLMMs in IPMs, and it should be possible to connect those year  
481 effects into a single model of multiple vital rates as we have done in the count-based  
482 projection matrix case.

483

484 *Conclusion* – Population modeling requires estimation of vital rates. Modeling vital rates  
485 with generalized linear mixed models is a powerful and flexible way to accomplish this.  
486 Here we've shown that generalized linear models that include random year effects in lieu of  
487 a specific explanatory variable can recover vital rate parameters and vital rate covariation  
488 well, and that a version of this approach with a YEAR effect shared by all GLMs of vital rates  
489 outperforms separate GLMs with respect to the estimation of vital rate covariation, when  
490 the exogenous effects are pervasive. The inclusion of a year effect in a model of vital rates  
491 is analogous to including a random block effect in an experimental setting. There, it is taken  
492 for granted that the random block effect improves parameter inference because it provides  
493 a better fit to the design of the data. We argue that models of vital rates should include  
494 random year effects (or they should at least be considered) for the same reason.

495

496 **Acknowledgements**

497

498 The authors thank Eric Menges, Pedro Quintana-Ascencio, and Eric Parent for useful  
499 discussions of *Dicerandra frutescens* demography and the models. MEKE was supported by  
500 the Agence Nationale de la Recherche, programme “6<sup>th</sup> Extinction”, through the project  
501 EVORANGE (ANR-09-PEXT-011).

502 **References**

503

504 Boyce, M. S. 1992. Population Viability Analysis. *Annual Review of Ecology and Systematics*  
505 23: 481-506.

506

507 Caswell, H. 2001. *Matrix Population Models, Second Edition*. Sinauer Associates, Inc.  
508 Sunderland, Massachusetts.

509

510 Carlin, B. P., J. S. Clark, and A. E. Gelfand. 2006. Elements of hierarchical Bayesian inference.  
511 In: Clark, J. S. and A. E. Gelfand (Eds.), *Hierarchical Modelling for the Environmental*  
512 *Sciences*. Oxford University Press, pp. 3-24.

513

514 Claus, M. J. and D. L. Venable. 2000. Seed germination in desert annuals: an empirical test  
515 of adaptive bet-hedging. *The American Naturalist* 155: 168-186.

516

517 Congdon, P. 2007. *Applied Bayesian Modelling*. Wiley Series in Probability and Statistics.

518

519 Coulson, T., J. M. Gaillard, and M. Festa-Bianchet. 2005. Decomposing the variation in  
520 population growth into growth contributions from multiple demographic rates. *Journal of*  
521 *Animal Ecology* 74: 789-801.

522

523 Coulson, T., E. A. Catchpole, S. D. Albon, B. J. T. Morgan, J. M. Pemberton, T. H. Clutton-  
524 Brock, M. J. Crawley, and B. T. Grenfell. 2001. Age, sex, density, winter weather, and  
525 population crashes in Soay sheep. *Science* 292: 1528-1531.



526

527 Doak, D., P. Karieva, and B. Klepetka. 1994. Modeling population viability for the desert  
528 tortoise in the western Mojave Desert. *Ecological Applications* 4:446-460.

529

530 Doak, D. F., W. F. Morris, C. Pfister, B. E. Kendall, and E. M. Bruna. 2005. Correctly  
531 estimating how environmental stochasticity influences fitness and population growth. *The*  
532 *American Naturalist* 166: E14-E21.

533

534 Easterling, M. R., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new  
535 structured population model. *Ecology* 81: 694–708.

536

537 Evans, M. E. K. and J. Dennehy. 2005. Germ banking: bet-hedging and variable release from  
538 egg and seed dormancy. *Quarterly Review of Biology* 80: 431-451.

539

540 Evans, M. E. K., K. E. Holsinger, and E. S. Menges. 2008. Modeling the effect of fire on  
541 demography of *Dicerandra frutescens* ssp. *frutescens*, an endangered plant endemic to  
542 Florida scrub. *Population Ecology* 50: 53-62.

543

544 Evans, M. E. K., K. E. Holsinger, and E. S. Menges. 2010. Fire, vital rates, and population  
545 viability: a hierarchical Bayesian analysis of the endangered Florida scrub mint. *Ecological*  
546 *Monographs* 80: 627-649.

547

548 Ezard, T. H. G., P. H. Becker, and T. Coulson. 2006. The contribution of age and sex to  
549 variation in common tern population growth rate. *Journal of Animal Ecology* 75: 1379-1386.

550

551 Gelfand, A. E., S. K. Sahu, and B. P. Carlin. 1995. Efficient parameterizations for normal  
552 linear mixed models. *Biometrika* 82:479-488.

553

554 Gelman, A. and D. B. Rubin. 1992. Inference from iterative simulation using multiple  
555 sequences. *Statistical Science* 7:457-511.

556

557 Gross, K., W. F. Morris, M. S. Wolosin, and D. F. Doak. 2006. Modeling vital rates improves  
558 estimation of population projection matrices. *Population Ecology* 48: 79-89.

559

560 Haridas, C. V. and S. Tuljapurkar. 2005. Elasticities in variable environments: properties and  
561 implications. *The American Naturalist* 166: 481-495.

562

563 Jongejans, E., H. de Kroon, S. Tuljapurkar, and K. Shea. 2010. Plant populations track rather  
564 than buffer climate fluctuations. *Ecology Letters* 13: 736-743.

565

566 Le Galliard, J. F., O. Marquis, and M. Massot. 2010. Cohort variation, climate effects, and  
567 populations dynamics in a short-lived lizard. *Journal of Animal Ecology*. 79: 1296-1307.

568

569 Menges, E.S. 1999. Ecology and conservation of Florida scrub. In: Anderson, R.C., J.S. Fralish,  
570 and J. Baskin (eds.), *The savanna, barren, and rock outcrop communities of North America*.  
571 Cambridge University Press, Cambridge, pp. 7-22.

572

573 Menges, E.S., P.F. Quintana-Ascencio, C.W. Weekley, and O.G. Gaoue. 2006. Population  
574 viability analysis and fire return intervals for an endemic Florida scrub mint. *Biological*  
575 *Conservation* 127:115-127.

576

577 Morris, W. F and D. F. Doak. 2002. *Quantitative Conservation Biology: Theory and Practice*  
578 *of Population Viability Analysis*. Sinauer Associates, Inc. Sunderland.

579

580 Morris, W. F., J. Altmann, D. K. Brockman, M. Cords, L. M. Fedigan, A. E. Pusey, T. S. Stoiski,  
581 A. N. Bronikowski, S. C. Alberts, and K. B. Strier. 2011. Low demographic variability in wild  
582 primate populations: fitness impacts of variation, covariation, and serial correlation in vital  
583 rates. *The American Naturalist* 177: E14-E28.

584

585 Orzack, S. H. and S. Tuljapurkar. 1989. Population dynamics in variable environments. VII.  
586 The demography and genetics of iteroparity. *The American Naturalist* 133: 901-923.

587

588 Sim, I. M. W., G. W. Rebecca, S. C. Ludwig, M. C. Grant, J. M. Reid. 2010. Characterizing  
589 demographic variation and contributions to population growth rate in a declining  
590 population. *Journal of Animal Ecology* 80: 159-170.

591

592 Stenseth, N. C., A. Mysterud, G. Ottersen, J. W. Hurrell, K.-S. Chan, and M. Lima. 2002.  
593 Ecological effects of climate fluctuations. *Science* 297: 1293-1296.

594

595 Tomimatsu, H. and M. Ohara. 2010. Demographic response of plant populations to habitat  
596 fragmentation and temporal environmental variability. *Oecologia* 162: 903-911.

597

598 Tuljapurkar, S. 1982. Population dynamics in variable environments. III. Evolutionary  
599 dynamics of r-selection. *Theoretical Population Biology* 21: 141-165.

600

601 Tuljapurkar, S. 1990. Population dynamics in variable environments. *Lecture Notes in*  
602 *Biomathematics* 185. Springer-Verlag, New York.

603

604 Tuljapurkar, S. and P. Weiner. 2000. Escape in time: stay young or age gracefully?  
605 *Ecological Modelling* 133: 143-159.

606

607 Tuljapurkar, S., J.M. Gaillard, and T. Coulson. 2009. From stochastic environments to life  
608 histories and back. *Philosophical Transactions of the Royal Society B* 364: 1499-1509.

609

610 Thomas, A., B. O'Hara, U. Ligges, and S. Sturtz. 2006. Making BUGS open. *R News* 6: 12-17.

611

612 Van Tienderen, P. H. 1995. Life cycle trade-offs in matrix population models. *Ecology* 76:  
613 2482-2489.

614

615 Venable, D. L. 2007. Bet hedging in a guild of desert annuals. *Ecology* 88: 1086-1090.

616 Table 1. Magnitude of the “true” effects of time-since-fire on vital rates (used for data simulation).

617

	<b>Profile 1</b>	<b>Profile 2</b>	<b>Profile 3</b>
$\beta_1^g$	-0.597	-0.597	-0.0597
$\beta_{1[1]}^{br}$	0.082	0	0
$\beta_{1[2]}^{br}$	-0.060	0	0
$\beta_{1[3]}^{br}$	-0.159	-0.159	0
$\beta_1^s$	0.780	0.780	0.0780
$\beta_1^{P1}$	-0.143	-0.143	0
$\beta_1^{P2}$	-0.195	0	0
$\beta_1^{P3}$	-0.599	-0.599	-0.0599
$\beta_1^{P4}$	-1.598	-1.598	-0.1598
$\beta_1^L$	-0.444	-0.444	0
$\beta_1^{Pr}$	0.142	0	0
$\beta_1^{SV}$	0.078	0	0
$\beta_1^{SS}$	-0.126	0	0
$\beta_1^{SM}$	0.052	0	0
$\beta_1^{SL}$	-0.200	-0.200	0
$\beta_1^{Rv1}$	-0.122	0	0
$\beta_1^{Rv2}$	0.119	0	0
$\beta_1^{Rv3}$	0.369	0.369	0
$\beta_1^{R1}$	-0.116	0	0
$\beta_1^{R2}$	-0.256	0	0

618

619

620 Table 2. Parameterization of three statistical models that estimate vital rates and their  
 621 covariation.  
 622

Model	TSF	YEAR effect	year effects	germination	initial seedling survival	seedling to vegetative transition
TSF	✓			$\beta_0^g + \beta_1^g \cdot \ln(TSF_{[t]})$	$\beta_0^s + \beta_1^s \cdot \ln(TSF_{[t]})$	$\beta_{0[3,2]}^{tr} + \beta_1^{P1} \cdot \ln(TSF_{[p,t]})$
YEAR		✓	✓	$\beta_0^g + \beta_2^g \cdot YEAR_{[t]} + \varepsilon_{yr[t]}^g$	$\beta_0^s + \beta_2^s \cdot YEAR_{[t]} + \varepsilon_{yr[t]}^s$	$\beta_{0[3,2]}^{tr} + \beta_{2[3,2]}^{tr} \cdot YEAR_{[t]} + \varepsilon_{yr[t,3,2]}^{tr}$
iYear			✓	$\beta_0^g + \varepsilon_{yr[t]}^g$	$\beta_0^s + \varepsilon_{yr[t]}^s$	$\beta_{0[3,2]}^{tr} + \varepsilon_{yr[t,3,2]}^{tr}$

623  
 624

625 **Figure Captions**

626

627 **Figure 1.** Estimates from the TSF model of (a) intercept terms for plant transitions ( $\beta_{0[i,j]}^{tr}$ )  
628 and (b) the effect of time-since-fire on “stasis” transitions (vegetative to vegetative, small  
629 flowering to small flowering, etc.). For each parameter, the true value (used for data  
630 simulation) is shown with a point. The plant classes are seedling (sd), vegetative (v), small  
631 flowering (s fl), medium flowering (m fl), and large flowering (l fl). Box limits fall at the first  
632 and third quartiles, whiskers end at 1.5 times the interquartile range, and open circles are  
633 outliers. Arrows highlight cases where the true value of the parameter falls outside the 95%  
634 CI of the estimated parameter value.

635

636 **Figure 2.** Difference from true time-since-fire effects on plant transitions (i.e., Profile 1) for  
637 three statistical models (TSF, YEAR, iYear). This includes all time steps and replicates for all  
638 plant transitions (15\*20\*25 values per statistical model). Box limits fall at the first and third  
639 quartiles, whiskers end at 1.5 times the interquartile range, and open circles are outliers.

640

641 **Figure 3.** Comparison between the true effect of time-since-fire (y-axis) and the effect, or  
642 its equivalent, estimated by three statistical models (x-axis) for: (a) initial seedling survival,  
643 (b) the transition from vegetative plant to large flowering plant, (c) small flowering plant to  
644 vegetative plant, and (d) large flowering to medium flowering plant. This comparison is  
645 made for each time step and replicate (20 \* 15 = 300 per statistical model) under Profile 1.  
646 The 1:1 line is shown in grey. The symbols are: red filled circles (TSF model), blue triangles  
647 (YEAR model), and green diamonds (iYear model).

648

649 **Figure 4.** Comparison of how well three statistical models (x-axis) estimate pairwise  
650 correlations between (a) the transition from seedling to vegetative plant and the transition  
651 from vegetative plant to small flowering plant, (b) initial seedling survival and the transition  
652 from seedling to vegetative plant, and (c) initial seedling survival and the transition from  
653 vegetative plant to vegetative plant, as simulated under parameter Profile 1. The  
654 correlation obtained from the simulated data is shown with a dashed horizontal line in each  
655 panel; zero is indicated with a solid horizontal line. Box limits fall at the first and third  
656 quartiles, whiskers end at 1.5 times the interquartile range, and open circles are outliers.  
657

658 **Figure 5.** Distribution of (a) bias and (b) root mean squared error of the estimates, from  
659 three statistical models (TSF, YEAR, iYear), of 79 pairwise correlations between vital rates,  
660 under three profiles of exogenous effects (detailed in Table 1). Bias is the average bias per  
661 pairwise combination of vital rates (across 15 replicate populations). Box limits fall at the  
662 first and third quartiles, whiskers end at 1.5 times the interquartile range, and open circles  
663 are outliers.