

Web Supplementary Materials

Models for Estimating Population Size from Repeated Counts of an Open Population

D. Dail^{1,*} and L. Madsen^{2,*}

^{*}Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

¹*email*: daild@science.oregonstate.edu

²*email*: madsenl@onid.orst.edu

Web Appendix A

Simplifying the Generalized Likelihood

With a T -fold infinite sum approximated by summing the first K terms of each sum, more than K^T terms would need to be summed to evaluate the likelihood in Equation (7) in the form presented in Section 3. Although this expression is useful for understanding how different model components contribute to the overall likelihood, it is not feasible to use in practice, especially with large T . Instead, (7) can be rearranged to take a recursive form that is much less computationally intense, requiring only T iterations of a summation over K terms.

To do this, first we let

$$\begin{aligned}g_1(N_{it}) &= \binom{N_{it}}{n_{it}} p^{n_{it}} (1-p)^{N_{it}-n_{it}} \\g_2(N_{i1}) &= \frac{e^{-\lambda} \lambda^{N_{i1}}}{N_{i1}!} \\g_3(N_{it}|N_{it-1}) &= P_{N_{it-1}, N_{it}}\end{aligned}$$

where $P_{N_{it-1}, N_{it}}$ is defined in Equation (6). So Equation (7) is then:

$$\begin{aligned}
& L(p_t, \lambda_i, \gamma, \omega \mid \{n_{it}\}) \\
&= \prod_{i=1}^R \left[\sum_{N_{i1}=n_{i1}}^{\infty} \cdots \sum_{N_{iT}=n_{iT}}^{\infty} \left\{ \left(\prod_{t=1}^T g_1(N_{it}) \right) \cdot g_2(N_{i1}) \cdot \prod_{t=2}^T g_3(N_{it} \mid N_{it-1}) \right\} \right]
\end{aligned}$$

Next we recursively define

$$\begin{aligned}
g^*(N_{iT-1}) &= \sum_{N_{iT}=0}^{\infty} g_1(N_{iT}) g_3(N_{iT} \mid N_{iT-1}) \\
g^*(N_{it-1}) &= \sum_{N_{it}=0}^{\infty} g_1(N_{it}) g_3(N_{it} \mid N_{it-1}) g^*(N_{it}), \text{ for } 2 \leq t \leq T-1
\end{aligned}$$

where each $g^*(N_{it-1})$ is a function of only N_{it-1} . Note that this recursive process goes in reverse-time, starting with $t = T-1$ and working backwards to $t = 2$. Finally, we have

$$L(p, \lambda, \gamma, \omega \mid \{n_{it}\}) = \prod_{i=1}^R \left\{ \sum_{N_{i1}=n_{i1}}^{\infty} g_1(N_{i1}) g_2(N_{i1}) g^*(N_{i1}) \right\}$$

which can be coded only using T infinite sums approximated by sums to K .

An adjustment is necessary for sites that are not sampled during $T = 1$. For any site i where this is the case, let t_i^* represent the first primary period that includes sampling at that site, and let Δ_{i1} represent the difference (in primary periods) between t_i^* and the first primary period. Then, we have:

$$L(p, \lambda, \gamma, \omega \mid \{n_{i1}, \dots, n_{iT}\}) = \sum_{N_{i1}=0}^{\infty} g_2(N_{i1}) \left[\sum_{N_{it_i^*}}^{\infty} g_1(N_{it_i^*}) g_3^{\Delta_{i1}}(N_{it_i^*} \mid N_{i1}) g^*(N_{it_i^*}) \right]$$

Assuming sites are independent allows us to then multiply the likelihood calculated for each site to get the overall likelihood.

Web Appendix B

Demonstration of Nested Likelihoods

We define

$$\Psi = \prod_{t=2}^T P_{N_{it-1}, N_{it}}$$

With $\omega = 1$ and $\gamma = 0$, we have $P_{N_{it-1}, N_{it}} = 0$ whenever $N_{it} \neq N_{it-1}$, and $P_{N_{it-1}, N_{it}} = 1$ only when $N_{it} = N_{it-1}$. So the generalized likelihood in Equation (7) is:

$$\begin{aligned} & L(p, \lambda, \gamma, \omega \mid \{n_{it}\}) \\ &= \prod_{i=1}^R \left[\sum_{N_{i1}=n_{i1}}^{\infty} \cdots \sum_{N_{iT}=n_{iT}}^{\infty} \left(\prod_{t=1}^T \text{Bin}(n_{it}; N_{it}, p) \right) \cdot \frac{e^{-\lambda} \lambda^{N_{i1}}}{N_{i1}!} \cdot \Psi \right] \end{aligned}$$

where

$$\Psi = \begin{cases} 1 & \text{if } N_{it} = N_{it-1} \text{ for all } 2 \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$$

which is equivalent to the likelihood given in Equation (2).

Web Table 1

Table 1: Summaries of the sampling distribution of abundance estimates using the generalized (Gen) and Royle models based on 1000 simulated data sets with $R = 20$ sites and $T = 5$ sampling occasions. The populations are subject to migration defined by γ and ω , whereas the Royle model assumes a closed population with $\omega = 1$ and $\gamma = 0$. q_1 , q_2 , and q_3 are the first, second, and third quartiles of $\hat{N}_{.5} - N_{.5}$. RMSE is root mean square error for each model. Coverage is the fraction of 95% “asymptotic” confidence intervals which contained the true value of $N_{.5}$. Open is the fraction of generalized models that, via AIC model selection, preferred an open population model over one of the Royle models.

γ	ω	λ	p	Model	$\hat{N}_{.5} - N_{.5}$			RMSE	Coverage	Open
					q_1	q_2	q_3			
0	1	2	0.25	Gen	-8.50	-0.37	10.06	17.15	0.986	0.004
				Royle	-8.29	-0.32	10.08	16.997	0.989	
0	1	5	0.25	Gen	-8.50	-0.37	10.06	17.15	0.986	0.004
				Royle	-8.29	-0.32	10.08	16.99	0.989	
0	1	2	0.50	Gen	-2.60	0.18	3.55	5.59	0.996	0.004
				Royle	-2.59	0.19	3.55	5.54	0.999	
0	1	5	0.50	Gen	-16.25	6.07	40.06	270.44	0.924	0.005
				Royle	-15.66	6.66	41.33	271.72	0.929	
1	0.8	2	0.25	Gen	-4.37	33.90	138.22	5875.24	0.909	0.070
				Royle	1.84	41.94	159.43	5878.25	0.958	
1	0.8	5	0.50	Gen	25.26	52.52	99.40	21143.32	0.945	0.081
				Royle	32.94	59.22	105.19	21143.35	0.998	
1	0.5	2	0.25	Gen	39.09	131.80	945.38	805.02	0.962	0.009
				Royle	40.06	134.19	946.08	856.35	0.968	
1	0.5	5	0.50	Gen	-11.48	75.59	169.22	17064.87	0.748	0.401
				Royle	99.23	156.13	294.88	74205.97	0.852	
2	0.8	2	0.25	Gen	-36.21	35.66	318.71	37380.31	0.869	0.216
				Royle	-8.24	63.58	360.41	37753.93	0.939	
2	0.8	5	0.50	Gen	-8.24	47.35	117.23	4563.51	0.909	0.207
				Royle	24.28	68.48	134.42	4577.24	0.992	
2	0.5	2	0.25	Gen	42.67	217.45	952.70	1959.62	0.873	0.117
				Royle	68.51	268.37	961.21	2082.12	0.926	
2	0.5	5	0.50	Gen	82.24	188.18	442.45	7181.55	0.833	0.135
				Royle	120.73	227.00	500.18	7290.64	0.918	

Web Table 2

Table 2: Summaries of the sampling distribution of abundance estimates using the generalized (Gen) and Royle models based on 1000 simulated data sets. The populations are subject to migration in a form other than expected by the generalized model, specified by the trend. The Royle model assumes a closed population with $\omega = 1$ and $\gamma = 0$. q_1 , q_2 , and q_3 are the first, second, and third quartiles of $\hat{N}_T - N_T$. RMSE is root mean square error for each model. Coverage (abbreviated “Cov.”) is the fraction of 95% “asymptotic” confidence intervals which contained the true value of N_T . Open is the fraction of generalized models that, via AIC model selection, preferred an open population model over one of the Royle models.

Trend	(R, T)	λ	p	Model	$\hat{N}_T - N_T$			RMSE	Cov.	Open
					q_1	q_2	q_3			
1.2	(100,10)	5	0.25	Gen	4.32	18.22	59.39	537.45	0.982	0.014
				Royle	4.65	18.70	61.08	538.53	0.994	
1.2	(100,10)	2	0.50	Gen	25.70	47.36	72.18	1171.17	0.523	0.196
				Royle	40.16	56.95	81.43	1279.32	0.511	
0.75	(20,5)	2	0.25	Gen	-864.70	-685.28	-381.33	21898.64	0.452	0.000
				Royle	-864.70	-685.28	-381.33	21898.64	0.452	
0.75	(20,5)	5	0.50	Gen	-152.08	-123.13	-91.99	129.48	0.689	0.000
				Royle	-152.08	-123.13	-91.99	129.48	0.689	

Web Table 3

Table 3: *Summaries of the sampling distribution of abundance estimates using the generalized (Gen) and Royle models based on 1000 simulated data sets. The models are compared using the same simulated data sets as reported in Web Table 2, but here the generalized model is allowed to take an additional form, where entering migration rate γ depends on previous site abundance, N_{it-1} . q_1 , q_2 , and q_3 are the first, second, and third quartiles of $\widehat{N}_T - N_T$. RMSE is root mean square error for each model. Coverage (abbreviated “Cov.”) is the fraction of 95% “asymptotic” confidence intervals which contained the true value of N_T . Open is the fraction of generalized models that, via AIC model selection, preferred an open population model over one of the Royle models.*

Trend	(R, T)	λ	p	Model	$\widehat{N}_T - N_T$			RMSE	Cov.	Open
					q_1	q_2	q_3			
1.2	(100,10)	5	0.25	Gen	4.32	18.22	59.39	537.45	0.982	0.014
				Royle	4.65	18.70	61.08	538.53	0.994	
1.2	(100,10)	2	0.50	Gen	25.70	47.36	72.18	1171.17	0.523	0.196
				Royle	40.16	56.95	81.43	1279.32	0.511	
0.75	(20,5)	2	0.25	Gen	-289.97	-126.68	-49.08	317.40	0.957	0.976
				Royle	-864.70	-685.28	-381.33	21898.64	0.452	
0.75	(20,5)	5	0.50	Gen	-65.55	-42.28	-19.65	55.52	0.895	1.000
				Royle	-152.08	-123.13	-91.99	129.48	0.689	

Web Table 4

Table 4: Estimated yearly route abundance of American robins from one route of the BBS using different migration models. p covariate is the covariate included for detection probability, where “Int” indicates only an intercept is fit, and “Lin” and “Fac” indicate sampling year is included as a linear term or as a factor, respectively. $f(N_{i1})$ is the prior distribution for initial site abundances, ΔAIC is the difference in AIC points between each model and the lowest AIC score observed, and \hat{N}_t is estimated route abundance during sampling of year t , with $t = 1$ representing calendar year 2003. Note that both the Royle and reshuffle models restrict the estimated total abundance to be constant across all sampling periods.

Model	$f(N_{i1})$	p covariate	ΔAIC	$\hat{N}_{.1}$	$\hat{N}_{.2}$	$\hat{N}_{.3}$	$\hat{N}_{.4}$	$\hat{N}_{.5}$	$\hat{N}_{.6}$
constant	Pois	Int	0	115.3	105.1	97.9	92.8	89.3	86.8
constant	Pois	Lin	0.81	157.1	97.9	69.0	55.0	48.1	44.8
Royle	Pois	Lin	0.91	192.5	192.5	192.5	192.5	192.5	192.5
Royle	Pois	Int	1.28	198.4	198.4	198.4	198.4	198.4	198.4
constant	NB	Int	1.96	117.0	106.6	99.2	94.0	90.4	87.9
autoreg.	Pois	Int	2.37	194.7	182.8	171.6	161.1	151.2	141.9
constant	NB	Lin	2.50	174.7	104.9	71.5	55.6	48.0	44.4
Royle	NB	Lin	2.91	193.8	193.8	193.8	193.8	193.8	193.8
Royle	NB	Int	3.27	205.8	205.8	205.8	205.8	205.8	205.8
autoreg.	Pois	Lin	4.36	197.9	183.3	169.8	157.3	145.7	135.0
autoreg.	NB	Int	4.37	194.7	182.7	171.5	161.0	151.1	141.9
constant	Pois	Fac	4.49	133.6	90.6	67.9	55.9	49.6	46.3
autoreg.	NB	Lin	6.36	198.6	184.1	170.6	158.1	146.5	135.8
constant	NB	Fac	6.48	134.0	90.8	68.0	55.9	49.6	46.2
Royle	Pois	Fac	6.54	186.1	186.1	186.1	186.1	186.1	186.1
Royle	NB	Fac	8.54	186.2	186.2	186.2	186.2	186.2	186.2
autoreg.	Pois	Fac	9.97	190.3	176.7	164.0	152.3	141.4	131.3
autoreg.	NB	Fac	11.98	190.4	176.8	164.1	152.4	141.4	131.3
reshuffle	Pois	Int	18.55	198.4	198.4	198.4	198.4	198.4	198.4
reshuffle	Pois	Lin	18.81	397.5	397.5	397.5	397.5	397.5	397.5
reshuffle	NB	Int	20.55	80.2	80.2	80.2	80.2	80.2	80.2
reshuffle	NB	Lin	22.55	51.3	51.3	51.3	51.3	51.3	51.3
reshuffle	Pois	Fac	25.13	186.1	186.1	186.1	186.1	186.1	186.1
reshuffle	NB	Fac	27.13	73.3	73.3	73.3	73.3	73.3	73.3