

This are electronic appendices to the paper by Jost & Ellner 2000 Testing for predator dependence in predator-prey dynamics: a non-parametric approach. *Proc. R. Soc. Lond. B* **267**, 1611–1620.

Electronic appendices are refereed with the text. However, no attempt has been made to impose a uniform editorial style on the electronic appendices.

Appendix A Gradient estimation

The gradient estimation method here is a refinement of the one used in Ellner *et al.* (1997), and is described in more detail by Ellner & Seifu (2000). The time series is smoothed by fourth-order local polynomial regression (Fan & Gijbels, 1996), and the linear coefficient of the fitted local polynomial at each observation time provides an estimate of the gradient (i.e., the time derivative of population density). Because the measurement errors here are small, a small bandwidth H is preferable; we used $H = 0.5$ d in all instances. (The bandwidth enters the local regression model as follows: the fitted value at time t_i is obtained by weighted least squares polynomial regression of $N(t_j)$ or $P(t_j)$ on t_j with weights $w_{ij} = \exp(-((t_i - t_j)/H)^2)$.) Even with such a small H the smoothing causes some bias towards zero in gradient estimates (i.e., peaks and troughs are rounded off a bit). To reduce the bias, we

- apply the estimation procedure to a spline interpolation of the data, for which the exact gradient at any time can be computed numerically
- use nonparametric regression to estimate a bias-correcting curve (the mean exact gradient as a function of the estimated gradients)
- un-bias the original gradient estimates by applying the estimated bias-correcting curve.

The bias correction is quite small (at most about 10% for the data sets here) but in simulation studies it can visibly improve the accuracy of fitted rate equations (Ellner & Seifu, 2000).

The gradient is estimated from the whole time series, but in all analyses here the estimates at the first two and last two observation times were deleted. Local polynomial fits tend to “wag” at the “tails” of the data, because they are constrained on only one side and therefore can chase after measurement errors in initial and final data points. The region where this occurs can be identified by comparing the gradient estimates to those obtained from local linear regression, which wags less due to the smaller number

of fitted coefficients. On the data here, fourth-order and linear local regression estimates were nearly identical except at the first and last data points in each series. Trimming two values at each end was a safety precaution, that sacrificed only a small fraction of the data.

Appendix B Testing the conservation of mass assumption

Taking the term $Nf(N_\delta)$ in equation (2) as known (i.e. using the logistic equation derived in Figure 2), the prey equation rearranges to

$$g(N, P) = \frac{rN(1 - N_\delta/K) - dN/dt}{P}.$$

Evaluating this expression at sampling times t_i gives a time series of estimated prey mortality due to predation, $\hat{g}(t_i)$. The predator equation re-arranges to the form

$$h(N_\tau, P_\tau) = \frac{1}{P} \frac{dP}{dt}.$$

Given the abundances P and the estimated gradient dP/dt , we can then produce a time series $\hat{h}(t_i)$. We then estimate the functions $g(N, P)$ and $h(N_\tau, P_\tau)$ by nonparametric regression of $\hat{g}(t_i)$ and $\hat{h}(t_i)$, respectively, on (n_i, p_i) and $(n_{i,\tau}, p_{i,\tau})$ ($n_{i,\tau}$ is a shortcut for $N(t_i - \tau)$, estimated by interpolation). The linear conversion assumption (together with a constant predator mortality rate) is that $h(N, P) = eg(N, P) - \mu$.

In order to maintain independence of the prey and predator equation for this comparison, we estimated τ from the predator equation alone by finding the value of τ that yields the best prediction of $\frac{1}{P} \frac{dP}{dt}$ as a function of (N_τ, P_τ) . Very similar results were obtained from several very different prediction methods: thin-plate splines (Wahba, 1990) and kernel regression with a range of reasonable bandwidths (both fitted using the FUNFITS package, Nychka *et al.* 1998), and projection pursuit regression with smoothing spline ridge functions (Roosen & Hastie (1994), fitted using the ASP package from Statlib (lib.stat.cmu.edu)). The estimated delay times are $\tau = 0.3$ d for series 11a, $\tau = 0.6$ d for series 12a, and $\tau = 0.1$ d for series 14c. Only the results for series 14c were at all sensitive to the prediction method, with the optimum τ varying between 0 and 0.2d, because the goodness-of-fit was relatively constant over this range for any of the prediction methods.

Figure 3 shows the results from this way of assessing the linear conversion assumption, using thin-plate splines with smoothing parameter selected by GCV2 to estimate $h(N, P)$ and $g(N, P)$. For each data series there is an approximately linear relationship in the scatterplot of $h(n_i, p_i)$ versus $g(n_i, p_i)$, with a negative intercept corresponding to the

positive predator mortality rate. The scatter in the plots simply results from the fact that the fitted two-dimensional functions g and h are not identical since both were estimated from noisy data sets. We therefore interpret Figure 3 as supporting the linear conversion assumption, at least as an acceptable first approximation within the accuracy of the data. Note however that although series 11a and 14c are under the same growth conditions, fitting 14c with the time-delay estimated for 11a increases the nonlinear discrepancy between g and h .

Appendix C The linear optimization problem

Let $(t_i, n_i, p_i)_{1 \leq i \leq q}$ be the observed time series, and $n_{i,\delta} (= N(t_i - \delta))$, $p_{i,\delta}$ the interpolated delayed abundances. Based on equations (4) and with x_i , y_i being the estimated gradients of prey and predator dynamics respectively, we define

$$\begin{aligned}
Y &= \left(x_1 - r\left(1 - \frac{n_{1,\delta}}{K}\right)n_1, \dots, x_q - r\left(1 - \frac{n_{q,\delta}}{K}\right)n_q, y_1, \dots, y_q \right)^T \in \mathcal{R}^{2q} \\
\theta &= (\beta_1, \dots, \beta_{m+d}, \mu)^T \in \mathcal{R}^{m+d+1} \\
W &= \left(\begin{array}{ccc|ccc} (c_n^2 n_1)^{-1} & & & & & \\ & \ddots & & & & \\ & & & & 0 & \\ \hline 0 & & (c_n^2 n_q)^{-1} & & & \\ & & & & (c_p^2 p_1)^{-1} & 0 \\ & & 0 & & & \ddots \\ & & & & 0 & & (c_p^2 p_q)^{-1} \end{array} \right) \in \mathcal{R}^{2q, 2q} \\
M &= \left(\begin{array}{c|ccc|c} 0_{d,d} & 0 & \cdots & 0 & 0_{d,1} \\ \hline 0 & \alpha & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \alpha & 0 \\ \hline 0_{1,d} & 0 & \cdots & 0 & 0_{1,1} \end{array} \right) \in \mathcal{R}^{m+d+1, m+d+1}
\end{aligned}$$

with $c_n = 0.417$ and $c_p = 0.165$. Using the spline base (5) with $z = N$, N/P or N/P^m , we can write

$$\begin{aligned}
Y_i = \sum_j X_{ij} \theta_j &= \begin{cases} -\hat{g}(z_i) p_i & 1 \leq i \leq q \\ e \hat{g}(z_k) p_k - \mu p_k & q+1 \leq i \leq 2q \end{cases} \\
&= \begin{cases} -\sum_{j=1}^d \beta_j z_i^j p_i - \sum_{j=1}^m \beta_{j+d} (z_i - \nu_j)_+^d p_i & 1 \leq i \leq q \\ e \sum_{j=1}^d \beta_j z_k^j p_k + e \sum_{j=1}^m \beta_{j+d} (z_k - \nu_j)_+^d - \mu p_k & q+1 \leq i \leq 2q \end{cases}
\end{aligned}$$

with $k = i - q$. The matrix X can thus be written as

$$X = \left(\begin{array}{c|c} & 0 \\ & \vdots \\ -B_0 & 0 \\ \hline & -p_1 \\ eB_\tau & \vdots \\ & -p_q \end{array} \right) \in \mathcal{R}^{2(q-1), m+d+1} \quad \text{with}$$

$$B_\tau = \left(\begin{array}{ccc|ccc} n_{1,\tau}p_1 & \cdots & n_{1,\tau}^d p_1 & (n_{1,\tau} - \nu_1)_+^d p_1 & \cdots & (n_{1,\tau} - \nu_m)_+^d p_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ n_{q,\tau}p_q & \cdots & n_{q,\tau}^d p_q & (n_{q,\tau} - \nu_1)_+^d p_q & \cdots & (n_{q,\tau} - \nu_m)_+^d p_q \end{array} \right),$$

$B_\tau \in \mathcal{R}^{q, m+d}$. The delayed population densities are computed by cubic interpolation from the observed densities (note that since we fitted on a reduced dataset (see Appendix A) and constrained the delay to be below 1 day, no densities had to be extrapolated).

In this notation we have to optimize the problem

$$F(e, \tau, (m)) = \min_{\theta} [(Y - X\theta)^T W (Y - X\theta) + \theta^T M \theta]. \quad (\text{C.1})$$

For the interpolated functions to be almost monotonic we require that $\hat{g}(s_{i+1}) > \hat{g}(s_i)$ (for some series s_i spanning the range of arguments of \hat{g}). We choose as the s_i 's the knots plus the endpoints, which is sufficient for monotonicity when using linear splines ($d = 1$), but even with the quadratic splines that we used this will give a reasonable form of the reconstructed functional response. Replacing \hat{g} by equation (5) and requiring parameter μ to be positive leads to a matrix inequality $C\theta \geq 0$ with $C \in \mathcal{R}^{m+2, m+d+1}$ defined by

$$\forall i, \quad C_{ij} = \begin{cases} s_{i+1}^j - s_i^j & j \leq d \\ (s_{i+1} - \nu_{j-d})_+^d - (s_i - \nu_{j-d})_+^d & d < j \leq m + d \\ \delta_{i-2, j-d-1} \text{ (Kronecker function)} & i = m + 2 \text{ or } j = m + d + 1 \end{cases}$$

Minimization of equation (C.1) under the constraint $C\theta \geq 0$ is a standard quadratic programming problem.

This nonparametric reconstruction is implemented in R (a freeware implementation of the S-language). The quadratic programming problem was solved with the contributed library `quadprog` that is based on the standard LINPACK packages. The nonlinear optimization problems were solved with the standard simplex algorithm in R (code adapted from Press *et al.* 1992).

Appendix D Methods for selecting the value of α

Our first selection method was *k-fold cross validation*, KCV. In this method the data are divided into nonoverlapping “blocks” (after rearranging the data in increasing order of the value of the independent variable in the regression to avoid temporal autocorrelation between blocks). The goodness of fit at a particular value of α is determined by repeatedly fitting the model to reduced data sets with one block omitted, and then generating predicted values for the omitted block. The KCV fitting criterion is the weighted sum of squared errors over all data points in all blocks. *Ordinary cross validation* (OCV) uses blocks of size 1. The purpose of using larger blocks is to eliminate spuriously good fits due to correlations in the data: a data point is not really “omitted” if the remaining data set has a high probability of containing a nearly identical data point. Block sizes were varied from 10 to 15 data points, each time estimating α by using a standard univariate optimizer, and the final α was taken as the maximum of these estimates to avoid overfitting.

The second method was by generalized cross validation (GCV, Wahba 1990). GCV is an approximation to OCV, in the sense that the smoothing parameter selected by GCV converges to that selected by OCV in the large sample limit. Following Nychka et al. (1992) we used the modified GCV criterion GCV2, which slightly overpenalizes model complexity in order to reduce the chance of spuriously overfitting. The GCV2 criterion estimates the smoothing parameter as the value of α which minimizes the expression

$$GCV2 = \frac{MSE}{(1 - 2\rho/n)^2},$$

where n is the number of fitted data points, MSE is the mean squared error, and ρ is the trace of the smoothing matrix,

$$\rho = \text{tr}(X(X^T W X + M)^{-1} X^T W).$$

In the presence of correlated errors, the GCV2 criterion must be modified to

$$GCV2 = \frac{MSE}{(1 - 2\Phi\rho/n)^2}$$

(Altman, 1999), where Φ is a measure of the spatial autocorrelation in the regression errors. If correlated errors are suspected, Φ can be estimated by making a pilot fit to the data and estimating Φ from the spatial correlation function of the residuals from the pilot fit (Altman, 1999). For our pilot fit we used the Hassell-Varley II functional response and found little or no spatial autocorrelation in the residuals, so we used $\Phi = 1$ for all datasets. The risk in using a particular parametric model for the pilot is that if the model cannot capture all of the structure of the data, residuals will have correlations (even if

the errors do not) and Φ will be overestimated. For our data this did not occur, as the pilot fit led to Φ being set at the lowest possible value.

Appendix E Raw data

prey growth in isolation

=====

CC = 0.5

From Figure 2b in Veilleux 1976, see our Figure 2(a)

time (days) density (#ind./ml)

0.00	15.58
0.50	30.04
1.00	66.05
1.50	141.60
2.00	274.60
2.50	410.00
3.00	468.80
3.50	526.40
4.00	472.50
4.50	496.60
5.00	489.50
5.50	492.00
6.00	496.80
6.50	473.00

CC = 0.375

From Figure 2c in Veilleux 1976, see our Figure 2(e)

time (days) density (#ind./ml)

0.00	16.56
0.50	24.67
1.00	42.25
1.50	74.04
2.00	133.10
2.50	230.00
3.00	355.40
3.50	393.20
4.00	401.30
4.50	363.10
5.00	403.20
5.50	374.60
6.00	385.10
6.50	376.60

predator-prey interaction

=====

CC = 0.5

From Figure 11a in Veilleux 1976, see our Figure 1(a)

time (d) prey(#ind/ml) predator(#ind/ml)

0.00	15.65	5.76
0.50	53.57	9.05
1.00	73.34	17.26
1.50	93.93	41.97
2.00	115.40	55.97
2.50	76.57	74.91
3.00	32.83	62.52
3.50	23.74	27.04
4.00	56.70	18.77
4.50	86.37	31.11
5.00	121.00	58.31
5.50	71.48	73.13
6.00	55.78	63.21
6.50	31.84	52.46
7.00	26.87	40.07
7.50	53.24	27.67
8.00	65.59	26.00
8.50	81.23	24.32
9.00	143.90	21.00
9.50	237.90	33.35
10.00	276.60	64.67
10.50	222.20	94.34
11.00	137.20	103.40
11.50	46.45	82.74
12.00	27.46	65.40
12.50	41.46	51.35
13.00	44.73	28.24
13.50	88.42	23.27
14.00	105.70	38.09
14.50	155.20	14.97
15.00	205.50	24.84
15.50	312.70	49.56
16.00	213.70	75.93
16.50	163.40	104.00
17.00	85.78	106.40
17.50	48.64	100.60
18.00	44.49	84.08
18.50	63.44	45.30
19.00	71.66	35.37
19.50	127.70	35.35

20.00	206.90	41.10
20.50	309.90	52.62
21.00	156.50	120.20
21.50	63.30	112.80
22.00	77.29	92.14
22.50	45.11	65.72
23.00	57.45	33.54
23.50	69.80	21.14
24.00	121.70	17.82
24.50	185.20	26.04
25.00	175.30	65.61
25.50	139.00	76.30
26.00	77.11	96.07
26.50	57.29	68.84
27.00	54.79	54.79
27.50	75.38	35.80
28.00	87.73	32.48
28.50	136.40	24.21
29.00	290.60	35.73
29.50	345.80	55.50
30.00	271.60	93.41
30.50	156.10	117.30
31.00	71.10	95.02
31.50	43.86	85.92
32.00	30.64	82.60
32.50	35.56	66.08
33.00	52.03	63.58
33.50	37.99	37.99
34.00	62.71	25.60
34.50	103.90	23.10
35.00	187.20	37.09

CC = 0.375

From Figure 12a in Veilleux 1976, see our Figure 1(b)

time (d) prey(#ind/ml) predator(#ind/ml)

0	61.21	3.448
0.5	86.21	3.448
1	94.83	5.172
1.5	138.8	9.483
2	212.9	9.483
2.5	302.6	10.34
3	346.6	14.66
3.5	311.2	24.14
4	212.1	38.79
4.5	162.9	63.79
5	93.1	91.38
5.5	63.79	78.45
6	67.24	47.41
6.5	80.17	30.17
7	87.93	14.66
7.5	106	18.1
8	131	20.69
8.5	175.9	24.14
9	223.3	19.83
9.5	274.1	23.28
10	283.6	27.59
10.5	233.6	41.38
11	180.2	55.17
11.5	162.1	69.83
12	125	57.76
12.5	105.2	51.72
13	111.2	38.79
13.5	128.4	31.03
14	113.8	28.45
14.5	129.3	16.38
15	133.6	8.621
15.5	149.1	12.93
16	182.8	18.1
16.5	200	19.83
17	225	27.59
17.5	201.7	34.48
18	176.7	57.76
18.5	150.9	42.24
19	129.3	34.48
19.5	124.1	11.21
20	132.8	13.79
20.5	161.2	15.52

21	203.4	20.69
21.5	235.3	25
22	212.9	37.93
22.5	184.5	50.86
23	150.9	53.45
23.5	137.1	51.72
24	123.3	38.79
24.5	126.7	35.34
25	131.9	25.86
25.5	156	10.34
26	184.5	15.52
26.5	240.5	25
27	256	39.66
27.5	219.8	56.9
28	220.7	45.69
28.5	193.1	22.41
29	150	9.483
29.5	161.2	6.897
30	179	8.621
30.5	196.6	10.34
31	223.3	16.38
31.5	243.1	28.45
32	254.3	43.97

CC = 0.5

From Figure 14c in Veilleux 1976, see our Figure 1(c)

time (d) prey(#ind/ml) predator(#ind/ml)

0	15.49	5.424
0.5	20.86	7.12
1	92.11	23.46
1.5	63.62	41.65
2	47.93	60.74
2.5	61.53	55.13
3	64.15	50.42
3.5	88.73	42.04
4	134.4	45.58
4.5	118.7	81.15
5	76.45	68.2
5.5	35.14	41.55
6	46.9	26.77
6.5	79.73	19.31
7	142.8	15.53
7.5	237.8	24.54
8	321.9	28.08
8.5	384.9	37.11
9	420.5	44.29
9.5	453.3	52.41
10	448.6	64.18
10.5	420.1	91.5
11	302.8	108.8
11.5	126	100.4
12	115.8	70.98
12.5	126.7	53.47
13	165.9	43.28
13.5	231.7	45.89
14	322.2	52.17
14.5	339.5	74.01
15	291.7	80.3
15.5	221.1	83.82
16	162.4	80.05
16.5	153.1	45.13
17	171.3	41.33
17.5	238.9	41.21
18	281.8	53.9
18.5	294.5	65.66
19	248.6	79.26
19.5	200.9	88.28
20	176	85.41
20.5	180.5	61.49

21	192.3	47.63
21.5	260.8	53.91
22	320.2	58.37
22.5	329.2	70.14
23	293.4	81.92
23.5	255.7	76.29
24	181.4	76.15
24.5	170.3	53.16
25	189.4	44.77
25.5	243.3	49.22

References

- Altman, N. S. 1999 Krige, smooth, both, or neither. Tech. rep., Department of Biometrics, Cornell University.
- Ellner, S. P., Kendall, B. E., Wood, S. N., McCauley, E. & Briggs, C. J. 1997 Inferring mechanisms from time-series data: delay-differential equations. *Physica D* **110**, 182–194.
- Ellner, S. P. & Seifu, Y. (2000) Fitting population dynamic models to time series data by gradient matching. (submitted).
- Fan, J. & Gijbels, I. 1996 *Local polynomial modelling and its applications*, vol. 66 of *Monographs on statistics and applied probability*. Chapman & Hall, London.
- Nychka, D. W., Haaland, P., O’Connell, M. & Ellner, S. P. 1998 FUNFITS, data analysis and statistical tools for estimating functions. In *Case Studies in Environmental Statistics*, eds. D. W. Nychka, W. W. Piegorsch & L. H. Cox, vol. 132 of *Lecture Notes in Statistics*, pages 159–179. Springer-Verlag, NY.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2nd edn.
- Roosen, C. B. & Hastie, T. J. 1994 Automatic smoothing spline projection pursuit. *Journal of Computational and Graphical Statistics* **3**, 235–248.
- Wahba, G. 1990 *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia PA.