Oikos

OIK-00814

Goyert, H. F., Manne, L. L. and Veit, R. R. 2014. Facilitative interactions among the pelagic community of temperate migratory terns, tunas and dolphins. – Oikos doi: 10.1111/oik.00814

Appendix 1

Details on methods of preliminary analyses, model structure and implementation

Preliminary analyses are useful in determining how to treat explanatory variables during model development, for establishing a framework around the stated hypotheses, rather than indiscriminately testing all possible combinations of covariates (data mining). We conducted preliminary analyses using the package MCMCglmm (Hadfield 2010), with an offset, random effect (date), and zero-inflated Poisson distribution. The lack of an autocorrelation term in this generalized linear mixed model (GLMM) means that it gives a reasonable estimate of parameter effects, while unveiling non-significant variables through its inflation of type I errors (Hawkins et al. 2007). SST gradient over 1 km (frontal strength) emerged as insignificant; hence, we did not consider it in the final Bayesian hierarchical models. For the zero-inflated negative binomial (ZINB) intrinsic conditional autoregressive (CAR) GLMMs, we conducted model selection across both seasons, combined (in addition to separated), to ensure that we were not abandoning a critical variable by discarding fluorescence from "fall" data: this dataset included fluorescence, and excluded observations that contained missing values (Table A3, models 3a-b). The ensuing ZINB CAR GLMM fit poorly to the mere 6208 observations, subset from 17 217, which amounted to less than half of the total n = 12 427 for the final 'spring' and 'fall' models. The survey effort for those final models corresponded to roughly 200 'spring' hours (n = 3712) and over 450 'fall' hours (n = 8715, for which there were more surveys but shorter daylight).

A general assumption in statistics is that variables are independent and identically distributed (i.i.d), yet rarely do ecological data abide by such terms. We corrected for violation in these assumptions by incorporating spatiotemporal model structure (to avoid type I errors), accounting for collinearity (to reduce type II error rates, Zuur et al. 2010), and using Bayesian inference to estimate the posterior distributions of parameters. For temporal structure, we chose date as a random effect, since it is nested within surveys, aiming to 1) allow spatially-overlapping survey 'legs' to be treated as distinct, 2) accommodate overnight gaps in data transects, and 3) remove interdependence between neighboring points at short lags (Zuur et al. 2009). We used a car.normal spatial structure for the residual term, which requires three matrices: adjacency (neighbor identity), number (neighbor count), and unnormalized weights – calculated using the neighborhood distance of 4 km (Figure A5) in package 'spdep' (Besag et al. 1991, Thomas et al. 2004, Saracco et al. 2010, Bivand et al. 2012, Saracco et al. 2012). We accounted for collinearity by ensuring quantities of two or less in the generalized variance inflation factors (GVIF) of all covariates considered in initial models (Zuur et al. 2010). We also scaled covariates by subtracting their means and dividing by their standard deviation, to generate unrelated sample chains (Zuur et al. 2012). Our model priors followed a uniform distribution for π, the random effect (a_i), and spatial residual term (ε_i), and a normal distribution for the intercept (α) and regression parameters (β). We ran the final models in WinBUGS ver. 1.4.3 (a Windows program for Bayesian inference using Gibbs sampling, Lunn et al. 2000), assessing significance of covariates after 100 000 iterations (those where the 95% confidence interval did not span zero), storing every 100th realization (the thinning rate) to reduce auto-correlation in the three chains, and setting burn-in at 10 000 samples.

Table A1. 2006–2009 offshore survey dates. Two ecosystems monitoring (EcoMon) Surveys, each extending from the mid-Atlantic (Cape Hatteras) to the Gulf of Maine / George’s Bank, made up the ‘spring’ data. Five surveys made up the ‘fall’ data: one EcoMon and four Atlantic Herring Hydroacoustic Surveys (shaded gray). A survey ‘leg’ refers to a spatiotemporally-delimited track either west or east of the longitude 71°W (Buzzards Bay, MA, Fig. 1a). ‘Fall’ data did not include the 2009 EcoMon Survey due to their missing salinity values. Therefore, lack of a yearly replicate for the southwesterly leg of the ‘fall’ 2008 EcoMon gave reason for its exclusion as well, thereby restricting ‘fall’ analysis to New England waters (Fig. 1b). For analyzing tern behavioral associations, data incorporated all surveys, and excluded records without tern behaviors.

Cruise code	Survey	Season	Year	Start date	End date
DEL0615	Herring	‘Fall	2006	19-Sep	28-Sep
DEL0706	EcoMon	‘Spring’	2007	23-May	3-Jun
DEL0710	Herring	‘Fall’	2007	14-Oct	25-Oct
DEL0808	EcoMon	‘Fall’	2008	17-Aug	28-Aug
DEL0809	Herring	‘Fall’	2008	4-Sep	9-Oct
DEL0905	EcoMon	‘Spring’	2009	28-May	10-Jun
DEL0909	EcoMon	‘Fall’	2009	17-Aug	28-Aug
DEL0910	Herring	‘Fall’	2009	12-Sep	15-Oct

Table A2. Occurrence, count, and density of tern species across all offshore surveys. Occurrence indicates the number of groups observed, counts enumerate individuals, and density is calculated from counts divided by sampling unit area (0.3 km^2, on average). Due to co-occurrence of species within sampling units, pooled occurrence differs from the sum of species occurrences.

Species	Occurrence	Count	Density (count km^{-2})
Common	341	1567	5283.8
Roseate	32	141	470.0
Arctic	18	32	107.3
Unidentified	149	2672	9727.1
Pooled	509	4412	15588.1

Table A3. Offshore tern habitat and community model selection by season, and seasons combined. The best models selected from the candidate set are shown in bold, and indicate: the final selected models 1a (‘spring’) and 2a (‘fall’), the number of observations (n), number of model parameters (Param.), mean coefficient and significance of each covariate (* $p < 0.05$, shaded in gray), and the deviance information criterion (DIC) value, based on the negative binomial distribution (Zuur et al. 2012). Covariates for habitat (columns at left) include: sea surface temperature (SST), fluorescence (chlorophyll concentration, ‘Chl’), salinity (‘Sal’), depth, and distance to shore (‘Dist’). Community covariates (columns at right) include: tuna or dolphin density (‘Dolph’), and number of whales per km (‘Whl’). The ‘x’ indicates parameters that were not tested (i.e. not included in the data), the ‘-’ indicates covariates that were not considered as predictors in the models shown (i.e. they were not significant in more complex models). We used a backwards stepwise approach, where the weak effects of most covariates offered few comparable final models; e.g. the second-best models (1b and

2b) had non-significant predictors. Models 3a and 3b summarize analyses across both seasons, combined, for which the dataset included fluorescence, and thereby excluded 11 009 observations with missing data – we chose not to report on these models, since they are not representative of the complete dataset, and have relatively poor fit.

				Habitat	•		Covariates		•	Community		
Model	Seasons	n	Param.	SST	Chl	Sal	Depth	Dist	Tuna	Dolph	Whl	DIC
1a	**'Spring'**	**3712**	**3**	**0.93***	**-**	**-**	**0.90***	**-**	**-**	**–1.21***	**-**	**27 813**
1b	'Spring'	3712	3	0.64*	-	–2.95	-	-	-	–0.20*	-	42 636
2a	**'Fall'"**	**8715**	**2**	**-**	**x**	**-**	**-**	**–1.21***	**0.22***	**-**	**-**	**443 165**
2b	'Fall'	8715	3	-	x	-	-	–1.12*	0.20*	0.09	-	444 778
3a	**Combined**	**6208**	**3**	**-**	**-**	**–0.24***	**0.38***	**-**	**0.14***	**-**	**-**	**124 194**
3b	Combined	6208	3	0.73*	-	-	-	–0.74*	0.13*	-	-	127 114

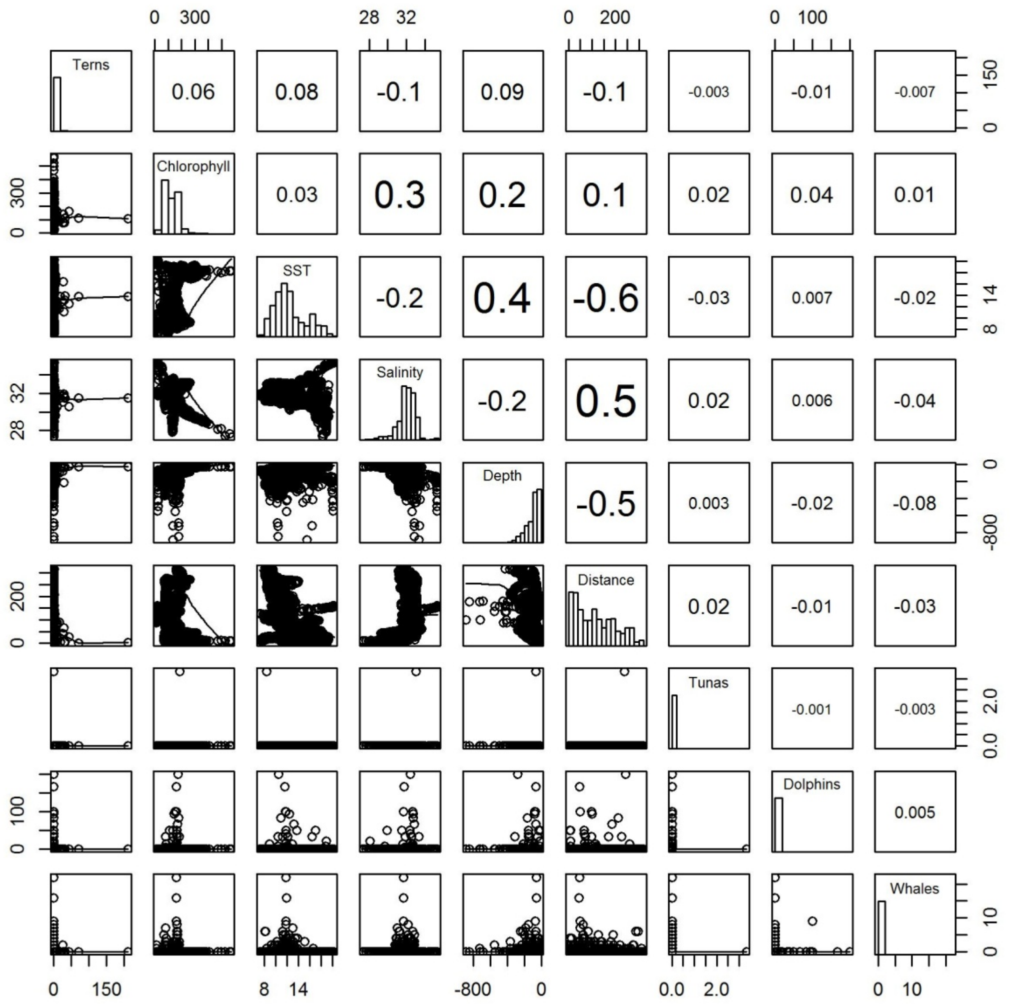


Figure A1 Relationships between terns and predictors in the 'spring' model. This pairplot (package 'AED', Zuur 2010) shows histograms of each variable on the diagonal, pair-wise Spearman rank correlation coefficients (font size proportional to value), and scatterplots with a LOESS smoother, axes labeled alternately (x at top or bottom, y at left or right). The response of pooled common and roseate terns, against the covariates considered for the "spring" model, is shown in the left column and top row (n = 3712). The variables considered as explanatory were: chlorophyll (fluorescence, parts per million), sea surface temperature (SST, °C), salinity (parts per thousand), depth (m), distance to shore (km), tuna density (km^{-2}), dolphin density (km^{-2}), and number of whales per km.

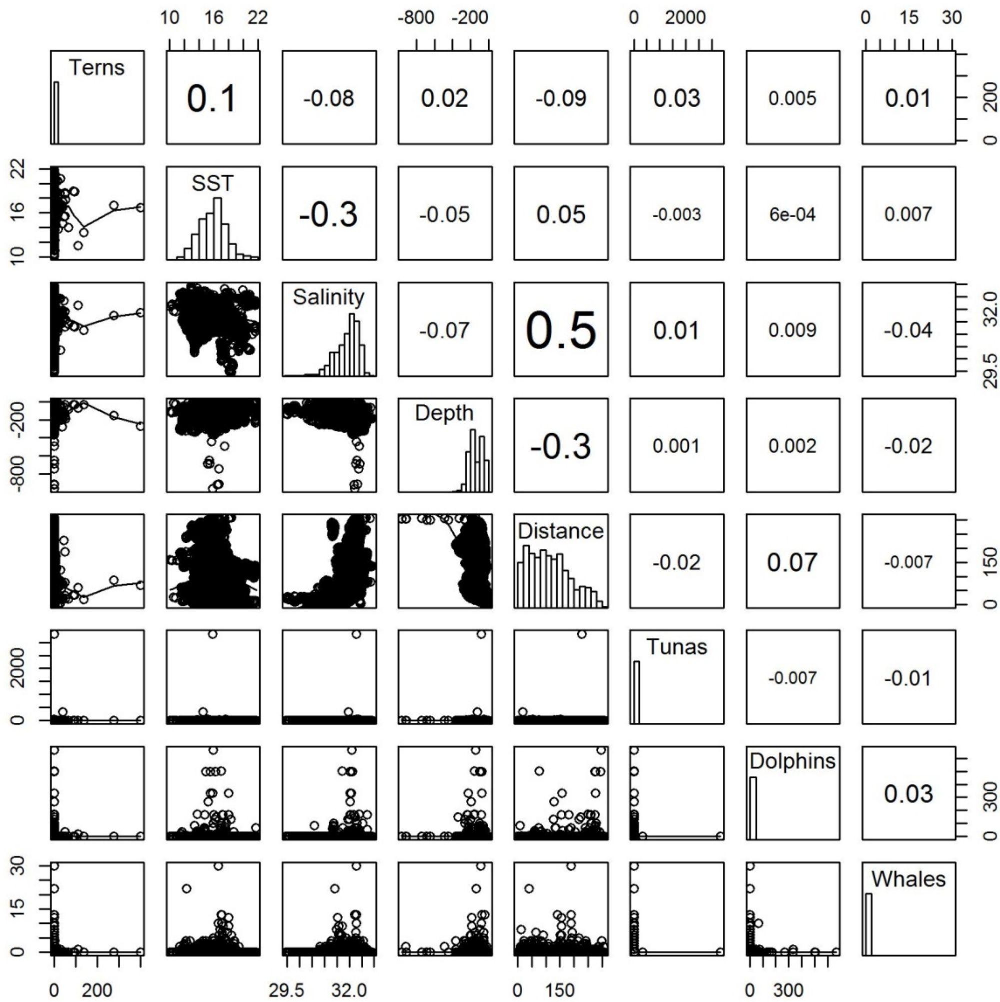


Figure A2 Relationships between terns and predictors in the ‘fall’ model. This pairplot (package ‘AED’, Zuur 2010) shows histograms of each variable on the diagonal, pair-wise Spearman rank correlation coefficients (font size proportional to value), and scatterplots with a LOESS smoother, axes labeled alternately (x at top or bottom, y at left or right). The response of pooled common and roseate terns, against the covariates considered for the “fall” model, is shown in the left column and top row (n = 8715). The variables considered as explanatory were: sea surface temperature (SST, °C), salinity (parts per thousand), depth (m), distance to shore (km), tuna density (km^{-2}), dolphin density (km^{-2}), and number of whales per km.

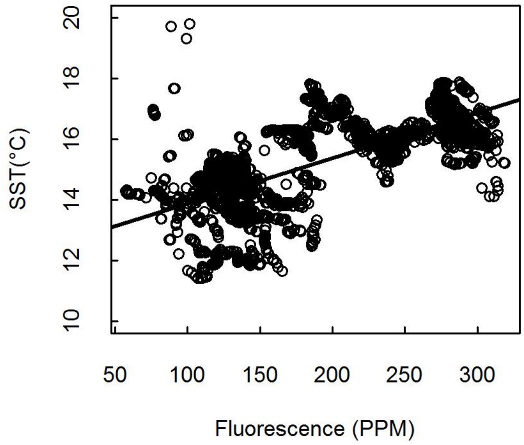


Figure A3. ‘Fall’ relationship between chlorophyll concentration and sea surface temperature (SST). We removed fluorescence (parts per million) from ‘fall’ data for modeling purposes, due to its high collinearity with SST, shown via regression (line slope = 0.015, R^2 = 0.47, $p < 1 \times 10^{-15}$).

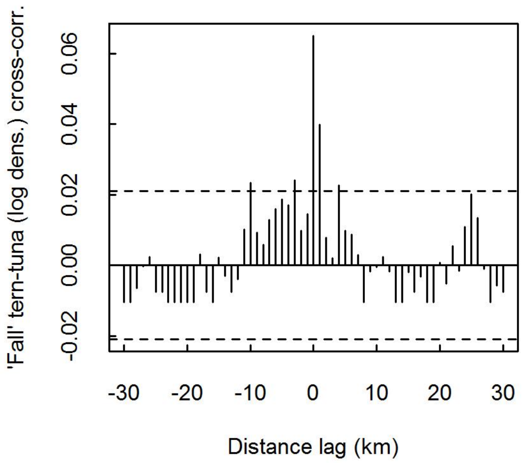


Figure A4. Spatial overlap between tern and tuna densities. The cross-correlation function (CCF), with 95% confidence intervals (dotted line), shows significantly cross-correlated 0–1 km lag distances (2 km patch sizes) of ‘fall’ log tern and tuna densities.

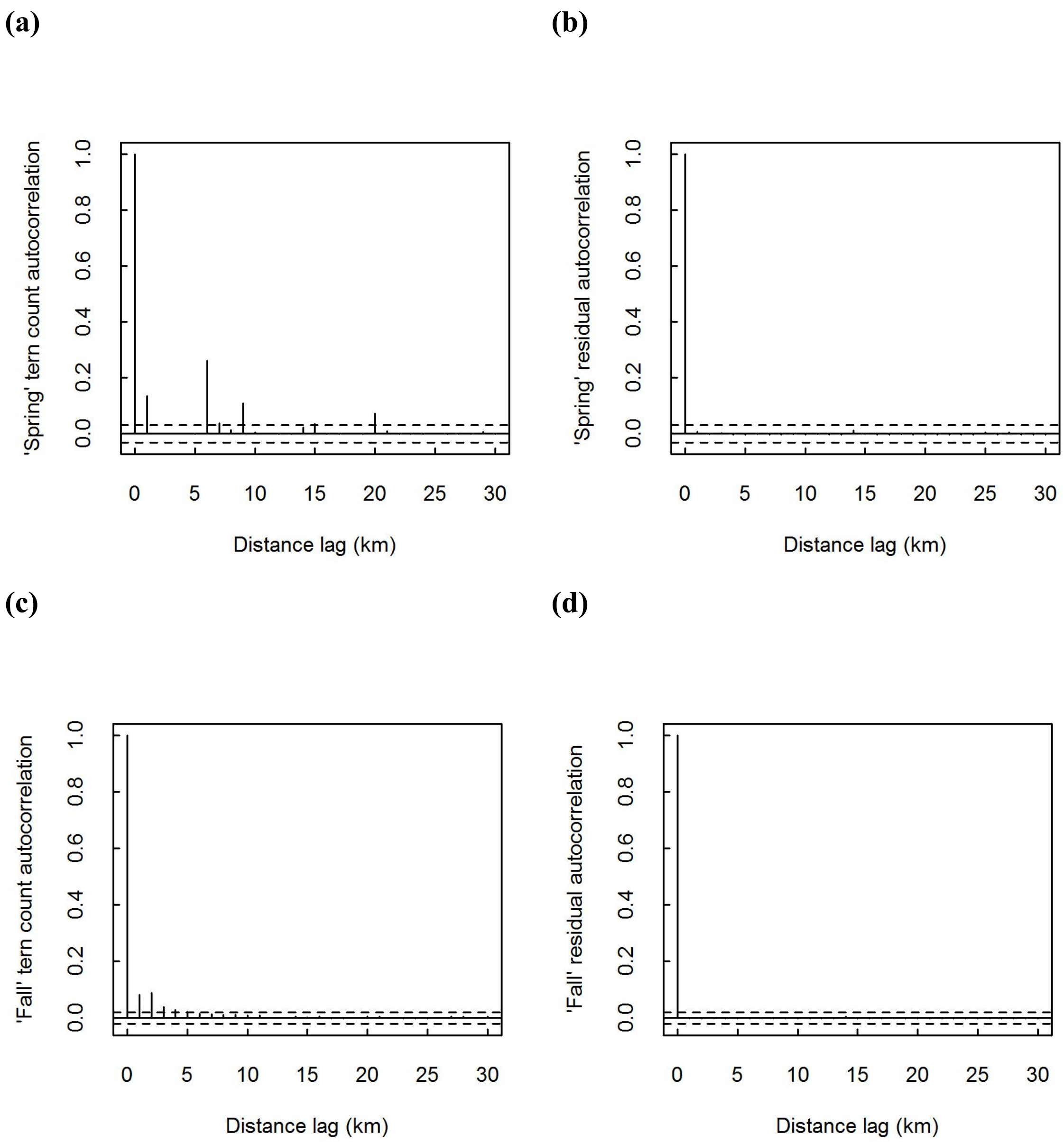
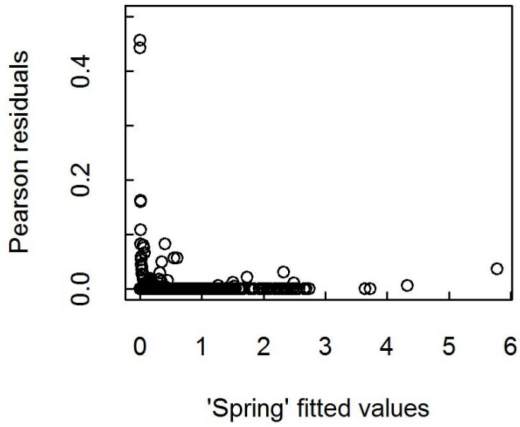
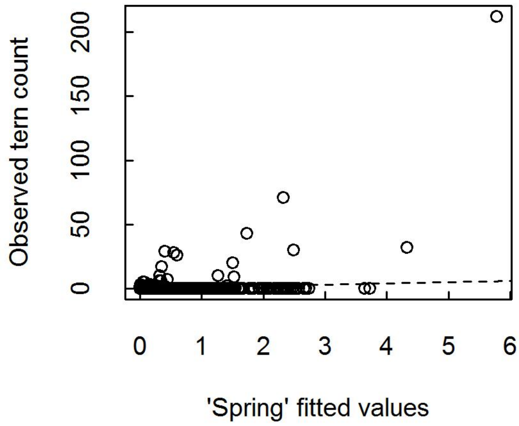


Figure A5. Spatial autocorrelation, as accommodated in the selected models. The autocorrelation function (ACF) with 95% confidence intervals (dotted line), shows significant autocorrelation at distance lags up to 2 km in ‘spring’ tern counts (a) and 4 km in ‘fall’ tern counts (c). The lack of autocorrelation in ‘spring’ (b) and ‘fall’ (d) Pearson residuals demonstrates that the model sufficiently accounted for autocorrelation in the response variable (no distance lags have a correlation value that is significantly greater than zero).

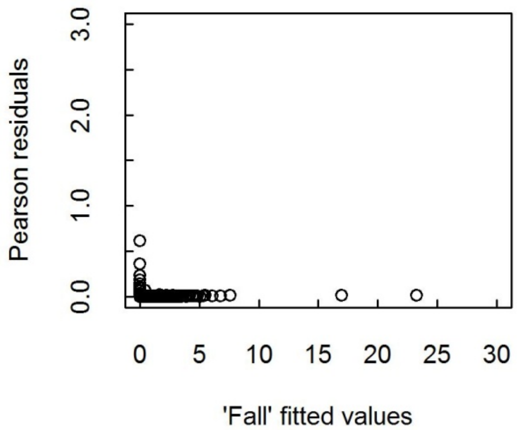
(a)



(b)



(c)



(d)

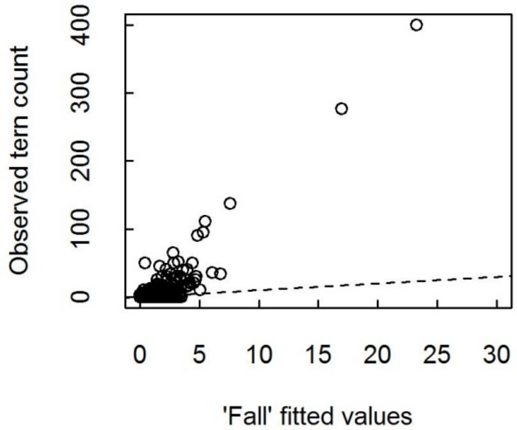


Figure A6. Relationship between values of the response variable, model fit, and residuals. Pearson residuals and observed tern counts are plotted against the models' fitted values for 'spring' (a, b) and 'fall' data, omitting the outlying fitted value of 4.80×10^{13} (c, d). The dotted line represents where observed and fitted tern counts would be equal, indicating model underestimation of the ecologically-important outliers (isolated large feeding aggregations).

References

Besag, J. et al. 1991. Bayesian image restoration, with two applications in spatial statistics. – Ann. Inst. Stat. Math. 43: 1–59.

Bivand, R. et al. 2012. spdep: spatial dependence: weighting schemes, statistics and models. – R package ver. 0.5-46.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. – J. Stat. Software 33: 1–22.

Hawkins, B. A. et al. 2007. Red herrings revisited: Spatial autocorrelation and parameter estimation in geographical ecology. – Ecography 30: 375–384.

Lunn, D. J. et al. 2000. WinBUGS – a Bayesian modelling framework: concepts, structure and extensibility. – Stat. Comput. 10: 325–337.

Saracco, J. et al. 2012. Spatial modeling of survival and residency and application to the monitoring avian productivity and survivorship program. – J. Ornithol. 152: 469–476.

Saracco, J. F. et al. 2010. Modeling spatial variation in avian survival and residency probabilities. – Ecology 91: 1885–1891.

Thomas, A. et al. 2004. Geobugs user manual. Ver. 1.2: <www.mrc-bsu.cam.ac.uk/bugs>.

Zuur, A. F. 2010. AED: Data files used in mixed effects models and extensions in ecology with R (Zuur et al. 2009). R package ver. 1.0.: <www.highstat.com/book2.htm>.

Zuur, A. F. et al. 2010. A protocol for data exploration to avoid common statistical problems. – Meth. Ecol. Evol. 1: 3–14.

Zuur, A. F. et al. 2009. Mixed effects models and extensions in ecology with R. – Springer.

Zuur, A. F. et al. 2012. Zero inflated models and generalized linear mixed models with R. – Highland Statistics, Ltd.