Juniper L. Simonis, Ethan P. White, S. K. Morgan Ernest. 2021. Evaluating Probabilistic Ecological Forecasts. *Ecology*.

APPENDIX S2

Additional details of the desert pocket mouse (*Chaetodipus penicillatus*) example.

Pocket Mouse Data and Summary Statistics

There are 24 50 m^2 (50 × 50 m) plots at Portal, each of which contains 49 permanent stations in a 7 × 7 grid that are sampled with Sherman live traps every lunar month. Four of the plots have always been available to rodents except for kangaroo rats, and the focal plot for the example is one of these four (plot 19). *C. penicillatus* has always been at the Portal site, but did not become prevalent in this plot until the 1990s, since when it has dominated the samples (Ernest et al. 2009, Ernest et al. 2016, Ernest et al. 2021). We accessed the data as version 2.80.0 on 2021-03-05 using R version 4.0.3 (R Core Team 2020) scripts (**Data S1**) leveraging version 0.3.6 of the portlar package (Yenni et al. 2020, Christensen et al. 2019).

We start our training data at sample 200 in the time series, corresponding to the date 1993-08-17, after which *C. penicillatus* has constituted 41.9% (729 of 1,740) of rodents trapped in the plot across the 290 complete surveys (out of 319 possible) through 2019-06-04 (Ernest et al. 2019). The next most abundant species during that time frame was 33.6% of the observations and all other species were less than 5% each (Ernest et al. 2019). Across those observations, *C. penicillatus* counts in the plot have cycled seasonally, ranging from 0 to 17 with a median of 1, a mean of 2.51, a variance of 8.77, and positive skew (skewness measures as 1.50 using the method of moments population estimate); the samples were 0-heavy (32.8%) and 45.9% of the samples contained 1 to 4 individuals (Fig. 4 in the main text; Ernest et al. 2019).

Fit and Analysis Details

Models were fit under a Bayesian framework via the Just Another Gibbs Sampler (JAGS, v.4.2.0; Plummer 2003, Plummer 2016) software, run from R (v4.0.3; R Core Team 2018) using the run.jags function in the runjags package (v2.0.4-6; Denwood 2016). Each model was fit using three separate chains, each of which was initialized at a random starting location then run for adaptation, burn-in, and sampling phases of 1,000, 5,000, and 10,000 steps, respectively. The 30,000 sampling steps were used without thinning to estimate parameters and the true count for each sample during the test period. We assessed chain convergence using autocorrelation, sample size adjusted for autocorrelation, and potential scale reduction factors (psrf, a.k.a. Gelman-Rubin statistic; Gelman and Rubin 1992).

Summary, analysis, and presentation were facilitated using custom R (v4.0.3; R Core Team 2020) scripts (**Data S1**). Portal data were accessed using the summarize_rodent_data function in the portalr package (v0.3.6; Christensen et al. 2019, Yenni et al. 2020). We processed the MCMC output using the as.mcmc.list, combine.mcmc, and as.mcmc functions in the coda package (v 0.19-4; Plummer et al. 2006). Calculation of the rank probability score was conducted via the crps_sample function in the scoringRules package (v1.0.1; Jordan et al. 2018, Jordan et al. 2020). We measured skewness of distributions using the skewness function in the e1071 package (v1.7-4; Meyer et al. 2020). The non-randomized PIT values were calculated using code based on that provided in Czado et al. (2009) (see **Data S1**).

Literature Cited

Christensen, E. M., G. M. Yenni, H. Ye, J. L. Simonis, E. K. Bledsoe, R. Diaz, S. D. Taylor, E. P. White, and S. K. M. Ernest. 2019. portalr: an R package for summarizing and using the Portal Project Data. *Journal of Open Source Software* **4**(33):1098.

Czado, C., T. Gneiting, and L. Held. 2009. Predictive model assessment for count data. *Biometrics* **65**:1254-1261.

Denwood, M. J. 2016. runjags: an R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software* **71**:1-25.

Ernest, S. K. M., T. J. Valone, and J. H. Brown. 2009. Long-term monitoring and experimental manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona, USA. *Ecology* **90**, 1708.

Ernest, S. K. M., G. M. Yenni, G. Allington, E. M. Christensen, K. Geluso, J. R. Goheen, M. R. Schutzenhofer, S. R. Supp, K. M. Thibault, J. H. Brown and T. J. Valone. 2016. Long-term monitoring and experimental manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona (1977–2013). *Ecology* **97**:1082.

Ernest, S. M., G. M. Yenni, G. Allington, E. Bledsoe, E. Christensen, R. Diaz, K. Geluso, J. R. Goheen, Q. Guo, E. Heske, D. Kelt, J. M. Meiners, J. Munger, C. Restrepo, D. A. Samson, M. R. Schutzenhofer, M. Skupski, S. R. Supp, K. Thibault, S. Taylor, E. White, D. W. Davidson, J. H. Brown, and T. J. Valone. 2018. The portal project: A long-term study of a Chihuahuan Desert ecosystem. bioRxiv. 10.1101/332783

Ernest, S. M., G. M. Yenni, G. Allington, E. Bledsoe, E. Christensen, R. Diaz, K. Geluso, J. R. Goheen, Q. Guo, E. Heske, D. Kelt, J. M. Meiners, J. Munger, C. Restrepo, D. A. Samson, M. R. Schutzenhofer, M. Skupski, S. R. Supp, K. Thibault, S. Taylor, E. White, D. W. Davidson, J. H. Brown, and T. J. Valone. 2021. Portal Project Data. v2.80.0. zenodo 10.5281/zenodo.3238678

Gelman, A. and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences.

Statistical Science **7**:457-511.

Jordan A., F. Krüger F, and S. Lerch. 2018. Evaluating Probabilistic Forecasts with scoringRules. *arXiv*. arXiv:1709.04743

Jordan, A., F. Krüger, and S. Lerch. 2020. scoringRules: Scoring rules for parametric and simulated distribution forecasts. R package version 1.0.1. https://CRAN.R-project.org/package=scoringRules.

Meyer, D. E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch. 2020. e1071: Misc. functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien. R package version 1.7-4. https://CRAN.R-project.org/package=e1071

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing,* K. Hornik, F. Leisch, and A. Zeileis, eds. ISSN 1609-395X.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**: 7-11.

Plummer, M. 2016. JAGS: Just Another Gibbs Sampler. v4.2.0. https://sourceforge.net/projects/mcmc-jags/files/JAGS/4.x/

R Core Team. 2018. R: A language and environment for statistical computing. v4.0.3. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Yenni, G. M., H. Ye, E. M. Christensen, J. L. Simonis, E. K. Bledsoe, R. M. Diaz, S. D. Taylor, E. P. White, and S. K. M. Ernest. 2020. portalr: Create useful summaries of the Portal data. R package version 0.3.6. https://CRAN.R-project.org/package=portalr.

Table S1. Component equations of the models used with the pocket mouse example.

Model name	Equations
Random Walk [RW]	$x_0 = \mu_0$ $\mu_n = x_{n-1}$ $\tau = \tau$ $x_n \sim \mathcal{N}(\mu_n, \tau)$
First-order autoregressive [AR(1)]	$x_0 = \mu_0$ $\mu_n = \varphi x_{n-1}$ $\tau = \tau$ $x_n \sim \mathcal{N}(\mu_n, \tau)$
Cyclic first-order autoregressive [cAR(1)]	$x_0 = \mu_0$ $\mu_n = \varphi x_{n-1} + \beta_1 \cos 2\pi j_n + \beta_2 \sin 2\pi j_n$ $\tau = \tau$ $x_n \sim \mathcal{N}(\mu_n, \tau)$

x: log-scale density, μ_0: log-scale density at time 0 (prior: Normal with mean $\log(\text{mean}(y))$, precision 0.25), $\mathcal{N}$ normal distribution (time varying mean μ_n and constant precision τ), φ: auto-regressive parameter (prior: Normal with mean 0, precision 1, and truncated at -1 and 1), β_1 and β_2: cyclic parameters (prior: Normal with mean 0, precision 0.16), j_n: fraction of the year at n, τ: precision (prior: Gamma with shape 1, rate 0.1). Samples n in 1 ... N are evenly spaced but an observation need not occur at every sample (allowing for missing observations).