Appendix B Data models used to estimate the probability of observations conditional on the predictions of the state-space model

In this appendix, we describe the probability distributions that link the observed data to the process model (main text, equation 1).

Total population size

The likelihood of the model's estimate of the total population size included sampling variance derived from multiple, annual censuses and a correction for animals that were not observed, p_{sight} (Table 3). We estimated observation error from the variances of replicate censuses occurring during the same June-July period each year. The average variance across all years ($var_{n.obs}$) was estimated from annually observed variances of multiple counts ($\mathbf{y}_{\text{N.var}}$),

$$[a_{var}, b_{var}|\mathbf{y}_{\text{n.var}}] \propto \prod_{t \in \mathbf{y}_{\text{n.varindex}}} \text{gamma}\left(y_{\text{n.var}(t)}|a_{var}, b_{var}\right) \text{gamma}\left(a_{var}|.001, .001\right) \text{gamma}\left(b_{var}|.001, .001\right)$$

$$var_{n.obs} = \left(\frac{a_{var}}{b_{var}}\right), \tag{B.1}$$

where $\mathbf{y}_{\text{n.varindex}}$ is a vector indexing years with non-missing observations for census variances. The likelihood for the estimate of the true population size $\eta_{(t)}$ was

$$[\mathbf{y}_{\text{n.obs}}, \boldsymbol{\eta}|var_{n.obs}, p_{sight}, \mathbf{N}] = \prod_{t=1}^{41} \text{Poisson}\left(y_{\text{n.obs}(t)}|\eta_{(t)} \cdot p_{sight}\right) \times \text{gamma}\left(\eta_{(t)} \;\middle|\; \frac{\left(\sum_{i=1}^{8} n_{(i,t)}\right)^2}{var_{n.obs}}, \frac{\left(\sum_{i=1}^{8} n_{(i,t)}\right)}{var_{n.obs}}\right). \tag{B.2}$$

Sex and age composition

The likelihood for the model's prediction of the proportion of the population in the juvenile class was

$$\left[\mathbf{y}_{\text{ratio.calf}}|\mathbf{n}_{(t)}, \mathbf{y}_{\sigma.\text{calf}}\right] = \prod_{t \in \mathbf{y}_{\text{i.ratio.calf}}} \text{beta}\left(y_{\text{ratio.calf}(t)} \;\middle|\; g\left(\frac{n_{(1,t)} + n_{(4,t)}}{\sum_{i=1}^{8} n_{(i,t)}}, y_{\sigma.\text{calf}(t)}\right)\right) \tag{B.3}$$

where $\mathbf{y}_{\text{i.ratio.calf}}$ is a vector indexing the years with non-missing data for aerial classifications, $y_{\text{ratio.calf}(t)}$ is the mean proportion of juveniles in groups observed during aerial counts at time t, $y_{\sigma.\text{calf}(t)}$ is the standard deviation of the mean, and $g(\quad)$ is the moment matching function that returns the parameters of a beta distribution given its mean and standard deviation. Procedures for estimating $y_{\text{ratio.calf}(t)}$ and $y_{\sigma.\text{calf}(t)}$ are given in appendix A.

We used data from ground counts to estimate the vector of parameters $\left(\boldsymbol{\alpha}_{\mu(t)}\right)$ of a Dirichlet distribution describing the proportions of the population contributed by non-juvenile males, adult females, yearling females, and juveniles. The total likelihood of the model predictions conditional on the data for these age classes was

$$\left[\mathbf{y}_{\text{p.}\mu(t)} \;\middle|\; \mathbf{n}_{(t)}\right] = \prod_{t \in y_{iground}} \text{Dirichlet}\left(\mathbf{y}_{p.\mu(t)} \;\middle|\; \left(\frac{n_{(8,t)}}{\sum_{i=1}^{8} n_{(i,t)}}, \frac{\left(n_{(3,t)} + n_{(6,t)} + n_{(7,t)}\right)}{\sum_{i=1}^{8} n_{(i,t)}}, \frac{\left(n_{(2,t)} + n_{(5,t)}\right)}{\sum_{i=1}^{8} n_{(i,t)}}, \frac{\left(n_{(1,t)} + n_{(4,t)}\right)}{\sum_{i=1}^{8} n_{(i,t)}}\right)' \sum_{j=1}^{4} \alpha_{\mu(i,t)}\right) \tag{B.4}$$

where $\mathbf{y}_{p.\mu(t)}$ is the vector of mean proportions of the population in each age class calculated as $\frac{\boldsymbol{\alpha}_{\mu(i,t)}}{\sum_{i=1}^{4} \alpha_{\mu(i,t)}}$. The parameters in the Dirichlet is a four element vector of the products of the sum of the parameters, multiplied by the model's estimate of the proportion of each age class in the population. Procedures for estimating $\boldsymbol{\alpha}_{\mu(t)}$ are given in appendix A.

Serology

Animals were classified as seropositive or seronegative as described in section *Seroprevalence*. Sensitivity and specificity of the serological test for exposure to brucellosis exceed 90% (Gall et al., 2000; Nielsen and Gall, 2001; Gall and Nielsen, 2004), but the uncertainty in test results should nonetheless be included in the data model for the model's estimates of exposure. We defined π as the probability that a test of a truly exposed animal gives a positive test result and ρ as the probability that a test of a truly un-exposed animal gives a positive test result. The likelihood for the model's estimate of exposure prevalence for the juvenile age class was

$$[\mathbf{y}_{\text{pos.calf}}|\mathbf{n}, \pi, \rho, \mathbf{y}_{\text{n.calf}}] = \prod_{t \in \mathbf{y}_{\text{isero.calf}}} \text{binomial}\left(y_{\text{pos.calf}(t)} \;\middle|\; \frac{\pi n_{(4,t)} + \rho n_{(1,t)}}{n_{(4,t)} + n_{(1,t)}}, y_{\text{n.calf}(t)}\right) \quad \text{(B.5)}$$

where $y_{\text{pos.calf(t)}}$ is the number of seropositive juveniles observed in a sample of size $y_{\text{n.calf}(t)}$ at time t and $\mathbf{y_{isero.calf}}$ is a vector indexing years with non-missing data. Likelihoods for exposure prevalence of juvenile and adult females followed the same form. Procedures for developing informative priors on beta distributions for π and ρ are given in section *Parameter models*.

Infectiousness

Capture histories (section *Infectiousness*) were used to estimate a time series of probabilities ($\mathbf{y}_\phi$) that a susceptible animal could become exposed via horizontal transmission during a year, which were used in out-of-sample validation (section *Model selection*). We estimated the unobserved, true state of the animal ($z_{(i,t)} = 1$ if exposed or $z_{(i,t)} = 0$ if susceptible at time t) and the probability of exposure $\left(y_{\phi(t)}\right)$ from observations of changes in serological status ($y_{mr(i,t)} = 1$ if

seropositive, $y_{mr(i,t)} = 0$ if seronegative) using

$$\left[\mathbf{z}, \mathbf{y}_{\phi} | \mathbf{Y}_{\mathrm{mr}}\right] \propto \prod_{i=1}^{79} \prod_{t_i \in y_{\mathrm{mr.index}(i)}} \mathrm{Bernoulli}\left(y_{mr(i,t)} | z_{(i,t)}\right) \prod_{i=1}^{79} \prod_{t=t_i+1}^{T_i} \mathrm{Bernoulli}\left(z_{(i,t)} | z_{(i,t-1)} \cdot y_{\phi(t)}\right) \times \prod_{t=1}^{14} \mathrm{beta}\left(y_{\phi(t)} | 1, 1\right) \quad \text{(B.6)}$$

where $\mathbf{Y}_{\mathrm{mr}}$ is the mark-recapture data set described above (section *Infectiousness*), $y_{\mathrm{mr.index}(i)}$ is the subset of years for which serological status was observed for animal i, t_i is the year of entry into the study of animal i, and T_i is the year of exit. Because we observed each animal's conversion from seronegative to seropositive during the year that the conversion occurred, we could assume that these animals would be infectious during their next pregnancy.

Literature Cited

Gall, D. and K. Nielsen. 2004. Serological diagnosis of bovine brucellosis: a review of test performance and cost comparison. *Review of Science and Techology* **23**:989–1002.

Gall, D., K. Nielsen, L. Forbes, D. Davis, P. Elzer, S. Olsen, S. Balsevicius, L. Kelly, P. Smith, S. Tan, et al. 2000. Validation of the fluorescence polarization assay and comparison to other serological assays for the detection of serum antibodies to brucella abortus in bison. *Journal of Wildlife Diseases* **36**:469–476.

Nielsen, K. and D. Gall. 2001. Fluorescence polarization assay for the diagnosis of brucellosis: A review. *Journal of Immunoassay & Immunochemistry* **22**:183–201.