Supplementary data 4. Model description and parameter estimates of the community occupancy model applied to small carnivore camera trapping data from a rural matrix.

Camera trapping was done in two villages, Vyeboom and Ka-Ndengeza, in the Vhembe District of South Africa. Villages were close to one another and we analyzed data as a single dataset. We used a Dorazio/Royle (DC) community occupancy model with data augmentation (DA) (Dorazio and Royle 2005) to analyse camera trapping data obtained from sampling the two villages, were we detected 9 small carnivore species and augmented the data with 13 potential species occurring in area (Table 1; Main Text). We specifically aimed to spatially estimate small carnivore species richness over the different land uses in order to investigate to potential ecosystem services that can be derived from small carnivore predation on pests. We followed a species specific parameterisation with random effects on detection and occupancy (e.g. species specific relationships with covariates).

The community occupancy model was parameterized as follows:

$$w_k \sim Bernoulli(\Omega) \text{ \# Superpopulation process}$$

$$z_{ik}|w_k \sim Bernoulli(w_k \psi_k) \text{ \# State process (occurance)}$$

$$y_{ijk}|z_{ik} \sim Bernoulli(z_{ik} p_{ijk}) \text{ \# Observation process (detection)}$$

$$\text{\# models of species heterogeneity (Eq. S1) } logit(\psi_{ik}) = lpsi_k + \beta 1_k \text{Cat}_i + \beta 2_k \text{Dog}_i + \beta 3_k Livestock_i$$

$$\text{\# (Eq. S2) } logit(p_{ijk}) = lp_k + \beta.dlp_k * Date_{ij}$$

with

$$lpsi_k \sim Normal(\mu_{lspi}, \sigma^2_{lspi})$$

$$\beta1_i \sim Normal(u_{\beta1}, \sigma^2_{\beta1})$$

$$\beta2_i \sim Normal(\mu_{\beta2}, \sigma^2_{\beta2})$$

$$\beta3_i \sim Normal(\mu_{\beta3}, \sigma^2_{\beta3})$$

$$lp_k \sim Normal(\mu_{lp}, \sigma^2_{lp})$$

$$\beta.d1_i \sim Normal(\mu_{\beta.d}, \sigma^2_{\beta.d})$$

In this model parameterization we augmented the observed species with species never detected (but that we hypothesised could occur in the study area), which is represented by a Bernoulli random variable (*w*), which indicates that the species is part of the metacommunity studied (e.g. data augmentation variable; (Kéry and Royle 2015)). z_{ik} represents the true occupancy state where 0 indicates not occupied and 1 occupied for a species *k* at site *i*; ψ_{ik} represents the occupancy probability (between 0 and 1) for each species *k*; $lpsi_k$ is the logit-linear predictor intercept of occupancy probability, which is indexed by species (*k*). *β1* is the coefficient for the Relative abundance of cats (expressed as number of pictures/1000 camera trapping days), *β2* the coefficient for relative abundance of dogs and *β3* coefficient for relative abundance of livestock. Species specific intercepts and coefficients comes from Normal distributions with mean (μ_{lpsi}), and variance (σ^2_{lpsi}) for the community, mean (*μβ1-μβ3*) and variance ($\sigma^2\beta1$- $\sigma^2\beta3$) for coefficients. Similarly, y_{ik} are the species detections (1 being detected and 0 not detected) of *k* species at *i* sites; p_{ik} is the detection probabilities per species, lp_k is the logit-linear predictor intercept of detection probability, which is again indexed by species (*k*). *β.d* is the effect of Julian survey date on detection probability. The species specific detection intercepts were drawn from a Normal distribution with community mean (μ_{lp}) and variance (σ^2_{lp}) and for Julian date coefficient mean (*μβ.d*) and variance ($\sigma^2\beta.d$).
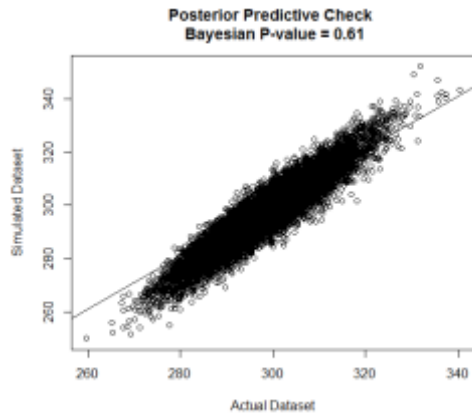
We used a Bayesian framework to implement the community occupancy model using JAGS (Plummer 2003) ported through R (RDevelopmentCoreTeam 2012) using the R package 'jagsUI' version 1.4.4 (Kellner 2015). We ran three parallel Markov chains with 50 000 iterations, where we disregarded 10 000 as burn-in and thinned the remaining chains by 10. We assessed chain convergence first by visually inspecting chains and calculating the Gelman-Rubin statistic (Gelman et al. 2014), where values of <1.1 indicated convergence. In our analysis all parameters had R values <1.1 which adequate chain mixing and convergence. We tested model fit by calculating the Bayesian p-value (Gelman et al. 1996) by comparing the observed residuals to residuals simulated under the model. Under perfect model fit we would expect the Bayesian p-value to be around 0.5, while values >0.95 indicate lack of fit.

We used the Freeman-Tukey residuals, R, in the calculation of the Bayesian p-value, where

$$R(\mathbf{y}, \boldsymbol{\theta}) = \sum (\sqrt{y} - \sqrt{E(y)})^2.$$

In this equation, $\mathbf{y}$ represent the binary observations, $\boldsymbol{\theta}$ represents all parameters in the community occupancy model. $E(y)$ is the expected value of y, which is the product of the species, site and the species specific detection and occupancy probabilities. The residuals are then summed over species, sites and occasions (see code for full parameterizing).

Our model simulations resulted in a Bayesian p-value of 0.61 which indicated a good fit of our community occupancy model.

Posterior Predictive Check
Bayesian P-value = 0.61

Simulated Dataset (y-axis) vs Actual Dataset (x-axis), axis values: 260, 280, 300, 320, 340

We present model parameter estimates only for the 9 detected species; for the augmented species model parameters are equivalent to the hyperparameter estimates. We report on the following parameters: (Table S3-1) is the community level parameters, (Table S3-2) species-specific estimates of the occupancy intercept *psi*, (Table S3-3) species specific estimates of *βcat*, (Table S3-4) species specific estimates of *βdog,* (Table S3-5) species specific estimates of *βlivestock*, (Table S3-6) species-specific estimates of the intercept for the logit-linear predictor of detection probability, (Table S3-7) and estimates of species-specific effect of Julian date on detection (on logit scale).

Table S3-1: Hyperparameter posterior summaries (metacommunity estimates) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior.

Parameter	Mean	SD	2.5%	50%	97.5%
Detection component					
μ_p	-3.020	1.135	-5.738	-2.744	-1.518
σ_p	1.513	0.783	0.321	1.423	2.906
μ_{date}	0.162	0.255	-0.380	0.162	0.673
σ_{date}	0.354	0.268	0.024	0.0295	1.036
Occupancy component					
μ_{psi}	-4.247	1.515	-7.587	-4.009	-1.982
σ_{psi}	2.383	1.025	0.814	2.247	4.586
$\mu_{\beta cat}$	-0.218	0.554	-1.726	-0.082	0.485
$\mu_{\beta dog}$	-1.898	0.491	-2.882	-1.892	-0.983
$\mu_{\beta livestock}$	0.913	0.349	0.227	0.920	1.593
$\sigma_{\beta cat}$	0.648	0.563	0.017	0.488	2.164
$\sigma_{\beta dog}$	0.385	0.325	0.019	0.305	1.199
$\sigma_{\beta livestock}$	0.351	0.309	0.012	0.272	1.164

Table S3-2: Posterior summaries of species-specific occupancy intercepts (psi in Eq. 1) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior.

Species	Mean	SE	2.5%	50%	97.5%
Civettictis civetta	-4.353	1.785	-7.656	-5.495	-0.723
Galerella sanguinea	-1.873	0.387	-2.764	-2.086	-1.221
Genetta maculata	-0.959	0.216	-1.396	-1.100	-0.547
Helogale parvula	-3.138	0.703	-4.448	-3.595	-1.637
Ichneumia albicauda	-2.474	0.381	-3.223	-2.728	-1.740
Ictonyx striatus	-3.590	1.532	-6.263	-4.618	-0.265
Mellivora capensis	-4.378	1.767	-7.706	-5.536	-0.696
Paracynictis selousi	-2.630	0.624	-3.848	-3.030	-1.368
Rhynchogale melleri	-4.359	0.804	-6.041	-4.848	-2.881

Table S3-3: Posterior summaries of the coefficient of cat relative abundance ($\beta 1$ in Eq. 1) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior. Bold font indicates strong effects with 95% Bayesian Credible Interval not overlapping 0.

Species	Mean	SE	2.5%	50%	97.5%
Civettictis civetta	-0.304	0.972	-2.927	-0.586	1.013
Galerella sanguinea	-0.653	0.959	-3.210	-1.024	0.373
Genetta maculata	0.253	0.256	-0.236	0.084	0.766
Helogale parvula	0.045	0.567	-1.267	-0.211	1.050
Ichneumia albicauda	0.189	0.340	-0.516	-0.015	0.849
Ictonyx striatus	-0.348	1.002	-3.003	-0.644	0.932
Mellivora capensis	-0.286	0.945	-2.823	-0.571	0.992
Paracynictis selousi	-0.570	1.014	-3.397	-0.902	0.502
Rhynchogale melleri	-0.365	0.933	-2.969	-0.639	0.731

Table S3-4: Posterior summaries of the coefficient of dog relative abundance ($\beta2$ in Eq. 1) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior. Bold font indicates strong effects with 95% Bayesian Credible Interval not overlapping 0.

Species	Mean	SE	2.5%	50%	97.5%
Civettictis civetta	**-1.941**	**0.662**	**-3.354**	**-2.329**	**-0.719**
Galerella sanguinea	**-1.949**	**0.537**	**-3.059**	**-2.292**	**-0.977**
Genetta maculata	**-1.975**	**0.486**	**-2.956**	**-2.294**	**-1.070**
Helogale parvula	**-1.894**	**0.599**	**-3.103**	**-2.274**	**-0.774**
Ichneumia albicauda	**-1.726**	**0.545**	**-2.776**	**-2.084**	**-0.662**
Ictonyx striatus	**-1.955**	**0.667**	**-3.403**	**-2.342**	**-0.751**
Mellivora capensis	**-1.944**	**0.678**	**-3.402**	**-2.330**	**-0.739**
Paracynictis selousi	**-1.849**	**0.578**	**-3.004**	**-2.224**	**-0.743**
Rhynchogale melleri	**-1.963**	**0.645**	**-3.368**	**-2.338**	**-0.791**

Table S3-5: Posterior summaries of the coefficient of livestock relative abundance ($\beta3$ in Eq. 1) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior. Bold font indicates strong effects with 95% Bayesian Credible Interval not overlapping 0.

Species	Mean	SE	2.5%	50%	97.5%
Civettictis civetta	0.827	0.566	-0.462	0.552	1.831
Galerella sanguinea	**0.932**	**0.363**	**0.229**	**0.693**	**1.664**
Genetta maculata	**0.809**	**0.340**	**0.135**	**0.583**	**1.465**
Helogale parvula	0.860	0.447	-0.086	0.592	1.706
Ichneumia albicauda	**1.118**	**0.426**	**0.369**	**0.833**	**2.060**
Ictonyx striatus	0.856	0.528	-0.317	0.581	1.846
Mellivora capensis	0.821	0.553	-0.425	0.552	1.800
Paracynictis selousi	**1.156**	**0.503**	**0.353**	**0.833**	**2.371**
Rhynchogale melleri	0.829	0.476	-0.229	0.571	1.678

Table S3-6: Posterior summaries for species-specific intercepts for the logit-linear predictor of detection probability, p, (as defined in Eq. 2) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior

Species	Mean	SE	2.5%	50%	97.5%
Civettictis civetta	-3.735	1.661	-7.309	-4.889	-1.293
Galerella sanguinea	-1.688	0.201	-2.091	-1.822	-1.312
Genetta maculata	-1.132	0.112	-1.357	-1.208	-0.917
Helogale parvula	-2.589	0.653	-4.030	-2.994	-1.474
Ichneumia albicauda	-1.956	0.339	-2.663	-2.178	-1.335
Ictonyx striatus	-3.798	1.432	-6.768	-4.804	-1.525
Mellivora capensis	-3.704	1.644	-7.212	-4.834	-1.276
Paracynictis selousi	-2.558	0.562	-3.695	-2.932	-1.498
Rhynchogale melleri	-1.941	0.599	-3.283	-2.292	-0.911

Table S3-7: Posterior summaries for species-specific effects of Julian date on detection probability on the logit scale (as defined in Eq. A2) for a community occupancy model fitted to data obtained from camera traps in a rural agricultural matrix, Vhembe District, South Africa, during 2014. Mean and SE = posterior mean and standard deviation; 2.5%, 50% and 97.5% = respective percentiles of the posterior

Species	Mean	SE	2.5%	50%	97.5%
Civettictis civetta	0.184	0.455	-0.772	-0.038	1.169
Galerella sanguinea	0.332	0.190	-0.011	0.196	0.714
Genetta maculata	-0.008	0.112	-0.230	-0.084	0.207
Helogale parvula	0.222	0.356	-0.501	0.020	0.974
Ichneumia albicauda	0.046	0.353	-0.789	-0.133	0.674
Ictonyx striatus	0.310	0.468	-0.537	0.045	1.395
Mellivora capensis	0.127	0.465	-0.925	-0.076	1.032
Paracynictis selousi	0.214	0.351	-0.546	0.022	0.913
Rhynchogale melleri	0.078	0.493	-1.089	-0.124	1.015

References

Dorazio, R. M. and J. A. Royle. 2005. Estimating size and composition of biological communities by modeling the occurrence of species. Journal of the American Statistical Association **100**:389-398.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. Bayesian data analysis. CRC press Boca Raton, FL.

Gelman, A., X.-L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica sinica:733-760.

Kellner, K. 2015. jagsUI: a wrapper around rjags to streamline JAGS analyses. R package version **1**.

Kéry, M. and J. A. Royle. 2015. Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models. Academic Press.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Page 125 *in* Proceedings of the 3rd international workshop on distributed statistical computing. Vienna, Austria.

RDevelopmentCoreTeam. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.