Matching observations and reality: using simulation models to improve monitoring under uncertainty in the Serengeti

Ana Nuno[1]*, Nils Bunnefeld[1,2] and E. J. Milner-Gulland[1]

[1]*Department of Life Sciences, Imperial College London, Silwood Park, Buckhurst Road, Ascot SL5 7PY, UK; and* [2]*School of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK*

Summary

1. Planning for conservation success requires identifying effective and efficient monitoring strategies but multiple types of uncertainty affect the accuracy and precision of wildlife abundance estimates. Observation uncertainty, a consequence of sampling effort and design as well as the process of observation, is still understudied, with little attention given to the multiple potential sources of error involved. To establish error minimization priorities and maximize monitoring efficiency, the direction and magnitude of multiple sources of uncertainty must be considered.

2. Using monitoring of two contrasting ungulate species in the Serengeti ecosystem as a case study, we developed a 'virtual ecologist' framework within which we carried out simulated tests of different monitoring strategies for different types of species. We investigated which components of monitoring should be prioritized to increase survey accuracy and precision and explored the robustness of population estimates under different budgetary scenarios.

3. The relative importance of each process affecting precision and accuracy varied according to the survey technique and biological characteristics of the species. While survey precision was mainly affected by population characteristics and sampling effort, the accuracy of the survey was greatly affected by observer effects, such as juvenile and herd detectability.

4. *Synthesis and applications*. Monitoring efficiency is of the utmost importance for conservation, especially in the context of limited budgets and other priorities. We provide insights into the likely effect of different types of observation and process error on population estimates for savanna ungulates, and more generally present a framework for evaluating monitoring programmes in a virtual environment. In highly aggregated species, the main focus should be on survey precision; sampling effort should be defined according to wildlife spatial distribution. For random or slightly aggregated species, accuracy is the key factor; this is most sensitive to observer effects which should be minimized by training and calibration by observer.

Key-words: bias, decision-making, monitoring errors, observation model, prioritization, savanna ungulates, survey methods, virtual ecologist

Introduction

The importance of ecological monitoring for conservation has often been acknowledged (Stem *et al.* 2005; Nichols & Williams 2006). Among its main objectives are to inform management decisions, measure success against stated objectives and learn about the system (Yoccoz, Nichols & Boulinier 2001). Monitoring is, however, often inadequate. Insufficient statistical power, lack of goal and hypothesis formulation, faulty survey design and data quality are common problems affecting monitoring schemes world-wide (Legg & Nagy 2006). The implications of these problematic issues are multiple; they not only affect monitoring effectiveness but also reduce resource availability for other potentially useful conservation interventions (McDonald-Madden *et al.* 2010). Resources for conservation are generally scarce (Bottrill *et al.* 2008), especially in developing countries

*Correspondence author. E-mail: ana.nuno08@imperial.ac.uk

(Danielsen *et al.* 2003). Planning for conservation success thus requires identifying effective and efficient monitoring strategies (Reynolds, Thompson & Russell 2011).

Monitoring is affected by multiple uncertainties (Harwood & Stokes 2003). Process uncertainty due to variation in the system itself (e.g. wildlife spatial distribution) interacts with observation uncertainty, which is a consequence of sampling effort and survey design as well as the process of observation. Observation uncertainty has multiple drivers and consequences. For example, estimates obtained from aerial surveys may be affected by a number of factors, such as animal detectability, observer performance, variation in aircraft height and deviations from the transect (Norton-Griffiths 1978; Jachmann 2002). Having imperfect knowledge of the true status of natural resources plays a central role in management decisions. For instance, Sethi *et al.* (2005) incorporated multiple types of uncertainty into a bioeconomic model of fisheries and found that observation uncertainty has the largest impact on policy, profits and extinction risk. The direction and magnitude of the effects of these processes on final abundance estimates have to be considered to establish error minimization priorities and maximize monitoring efficiency.

Optimization of sampling effort to achieve monitoring goals is demonstrably an essential consideration (Field, Tyre & Possingham 2005; Sims *et al.* 2006), but considerably less attention has been given to the effects and, particularly, the drivers of observation error. The effects of undercounting or the misidentification of the sex or age of an individual have received limited attention (Elphick 2008), most likely because multiple processes may occur simultaneously and discerning their impacts from monitoring data may be difficult. Knowing which types of errors are most important and should be tackled first is particularly challenging. Experimentation is often difficult, due to terrain, lack of capacity and the financial and time costs involved. For convenience and model simplicity, observation uncertainty is often considered as an overarching composite process when using simulations, modelled through log-normally distributed errors (e.g. Hilborn & Mangel 1997; Shea & Mangel 2001).

Modelling is a particularly useful tool because it allows experimentation through simulation. Previous studies have used modelling, for example, to investigate how to improve survey effort and design but without taking specific errors in the observation process into consideration (Sims *et al.* 2006; Blanchard, Maxwell & Jennings 2008), correct observation bias based on herd size detectability (McConville *et al.* 2009), assess the effects of data quality on harvest strategies and income (Milner-Gulland, Coulson & Clutton-Brock 2004), and estimate the risk of failing to detect a trend and wasting resources (Katzner, Milner-Gulland & Bragin 2007). Using a modelling approach, it is possible to explicitly simulate 'true' scenarios of wildlife abundance and distribution. Each step of the observation procedure can then be replicated to

investigate how the quality of the data collected ('observed state') may be improved and particularly how researchers' actions and assumptions affect precision (uncertainty or variability in the estimates which is used to produce confidence intervals around them) and accuracy (difference between the set of estimates and the truth they represent).

The Serengeti ecosystem is one of the most intensively studied systems in Africa. Long-term research in the Serengeti includes monitoring of a range of species, with wildlife censuses having been conducted since the 1950s (Sinclair *et al.* 2007). Monitoring resources are, however, very limited, especially given that this ecosystem covers more than 25 000 km^2. Monitoring must therefore be adjusted according to available budgets, while still being able to provide accurate and precise abundance estimates. Using monitoring of two contrasting ungulate species in the Serengeti ecosystem as a case study, we employed simulation modelling to investigate how abundance estimates are affected by multiple types of uncertainty, with a focus on observation error. Specifically, we investigated which factors should be prioritized to increase survey accuracy and precision and explored the potential effects of different budgetary scenarios on the robustness of the population estimates obtained for species of different ecological characteristics. This enables us to provide insights into the likely effect of different types of observation and process error on population estimates for savanna ungulates and more generally to present a framework for evaluating monitoring programmes in a virtual environment.

Materials and methods

STUDY AREA AND SPECIES

We chose two species to investigate the contrasting issues involved in monitoring ungulate species in savanna ecosystems. The migratory wildebeest *Connochaetes taurinus* population is monitored throughout the Serengeti ecosystem using aerial surveys to take photographs within sampling blocks (Fig. S1, Supporting Information). By contrast, a resident population of impala *Aepyceros melampus* is monitored using systematic flights along transects in a Game Reserve adjoining the Serengeti National Park (Grumeti–Ikorongo GR). Surveys are conducted approximately every 3–5 years in February/March to assess populations of resident and migratory ungulates (Campbell & Borner 1995; TAWIRI 2010).

The wildebeest population is highly gregarious and composed of bachelor herds and large nursery herds, with territorial males at certain times of year (Estes 1992). Wildebeest use the Serengeti plains in large herds during the wet season (mid-October through April), when the monitoring is conducted, moving west and north at the beginning of the dry season (May to mid-October). They give birth synchronously in February (Thirgood *et al.* 2004). Impala occur in the Serengeti woodlands and their populations are composed of large groups of females with a single dominant male (Jarman & Jarman 1973). These sedentary ungulates move up to 3 km in the dry season and 0·95 km in the wet season

(Estes 1992). Currently, there are around 1·3 million wildebeest in the Serengeti and 10 000 impala in the Game Reserve (Grumeti Fund 2010; Hilborn & Sinclair 2010).

METHODOLOGICAL FRAMEWORK

We simulated the monitoring process for the two ungulate populations to investigate monitoring precision and accuracy. The methodological framework was divided into four main components (Fig. 1): (a) a spatial distribution model, which provided the 'true scenario' against which simulated monitoring data were compared; (b) an 'observation model', which simulated monitoring of these populations; (c) a data analysis component, which estimated wildlife abundance from simulated monitoring data, the 'assessment model'; and (d) an assessment of survey accuracy and precision, in which discrepancies between 'true' and 'observed' population sizes and their drivers were investigated.

MODELLING THE DISTRIBUTION OF WILDLIFE

Wildebeest

A virtual wildebeest population was distributed in a 90×35 km grid with a total area of 3150 km^2 (2887 km^2 was surveyed in 2009; Hilborn & Sinclair 2010). Grid units were sized to be equivalent to a 'potential photograph' capturing around 0·05 km^2. Numbers of wildebeest per grid unit were simulated using a negative binomial distribution (NBD) with two defining parameters: the mean (μ) and the aggregation (k), with lower values of k representing more aggregated populations. Pieters *et al.* (1977) compare the efficiencies of several methods of estimation of the parameter k; the method of moments estimate relates it to the empirical mean μ and variance σ^2 by:

$$k = \frac{\mu^2}{\sigma^2 - \mu} \qquad \text{eqn 1}$$

A NBD allows us to account for, and investigate the effects of, differing degrees of animal aggregation on survey counts (Matthiopoulos 2011). To check its suitability to describe wildebeest counts, the goodness-of-fit was assessed through comparison of fitted and actual counts from the 2006 census ($\chi^2 = 338$, d. f. = 887, $P > 0.99$; J. G. C. Hopcraft, unpublished data).

The number of wildebeest in each cell is likely to be affected by abundance in neighbouring cells, so we adopted a geostatistical approach to incorporate spatial autocorrelation using the R package *geoR* (Ribeiro & Diggle 2001, version1·7-4). First, we defined the spatial autocorrelation structure by using a Gaussian process with variance–covariance matrix C related to an exponential correlation function between unit locations:

$$C = \sigma_s^2 \exp(-d_{ij}/\text{phi}) \qquad \text{eqn 2}$$

where d_{ij} is the distance between grid units i and j, σ_s^2 is the threshold variance known as the sill (which we kept at a constant value) and phi is the range parameter that represents a fraction of the distance beyond which there is little or no autocorrelation (Diggle, Tawn & Moyeed 1998). The strength of spatial autocorrelation was controlled by varying the range parameter; the larger the range, the stronger the autocorrelation because it persists over longer distances. To generate spatially autocorrelated survey counts, we then conditioned the outcome of the NBD on these spatially correlated random fields by affecting the actual realization of the distribution for each cell via the exponential link and the mean and aggregation parameters.

At the time of the counts, juvenile wildebeest are found within large nursery herds with their mothers, while older males remain in separate aggregations (Estes 1992). In the 2006 census in the
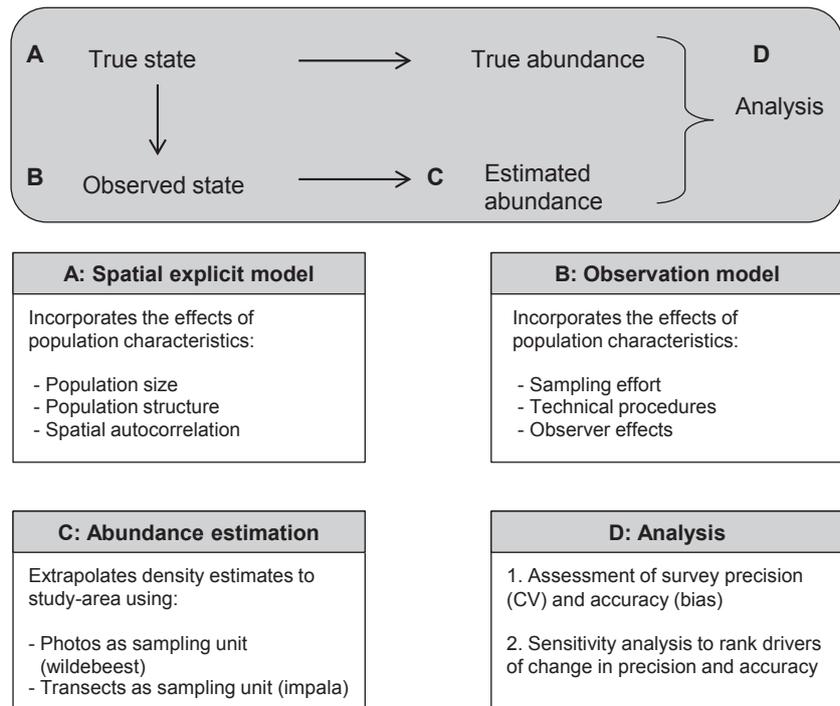

Fig. 1. Conceptual description of the study's methodological approach.

Serengeti, juvenile wildebeest were more likely to be present in photographs with higher total numbers of animals (Fig. 2; GLM with a binomial error structure: $z = 6.560$, d.f. = 340, $P < 0.001$). Empirical juvenile counts did not differ significantly from a NBD ($\chi^2 = 84$, d.f. = 340, $P > 0.99$), and 46% of the photographs with wildebeest present had juveniles and adults, while the remaining only had adults (J.G.C. Hopcraft, unpublished data).

The distribution of juveniles was modelled for a range of juvenile proportions in the total population. Juvenile counts followed a NBD drawn separately from the previous one which simulated the total count of wildebeest per cell, incorporating both adults and juveniles. We assumed that juveniles occurred in half of the photographs where wildebeest were found, with the probability of presence associated with higher total numbers of animals. Juveniles were thus redistributed according to total wildebeest counts per cell. To simulate the number of juveniles per cell, the desired proportion of cells without juveniles [i.e. zeros in the NBD; $p'(0)$] and mean number of juveniles per cell μ_{juv} were used to estimate the aggregation k_{juv} of juveniles (Perry & Taylor 1986):

$$p'(0) = \left(\frac{1 + \mu_{juv}}{k_{juv}}\right)^{-k_{juv}} \qquad \text{eqn 3}$$

Impala

A range of 'true' population sizes of impala was distributed in a virtual landscape with shape and area similar to the real survey area (around 1500 km^2; Fig. S1, Supporting Information).

Impala form herds of two to hundreds of animals and are generally dispersed in a random or slightly aggregated pattern (Jarman & Jarman 1973; Stein & Georgiadis 2008). We were interested in modelling individual spatial locations while taking into account the abundance patterns related to herd distribution and size. Impala distribution was thus modelled using a 3-step clustering process: (1) a number of clusters ('herds') was situated
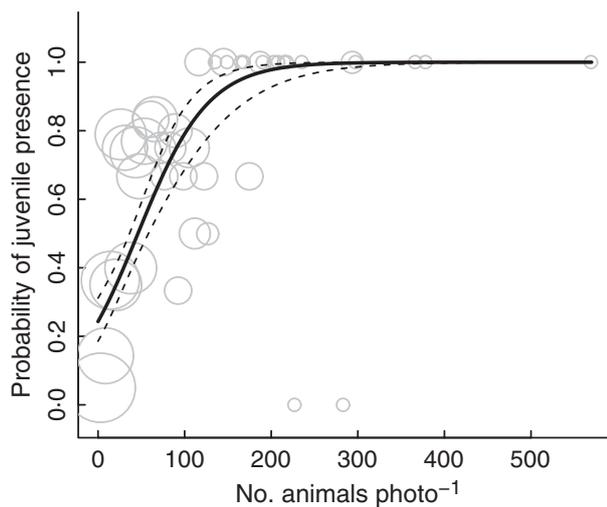
randomly in the landscape (assuming a homogeneous landscape) according to a Poisson process, defining a 'parent point' per herd; (2) herd size followed a log-normal distribution; and (3) animals within each herd were independently and uniformly scattered inside a circular herd home range centered on the parent point. Herd home range was obtained by multiplying herd size by the assumed 'individual space' requirements, up to a specified maximum value ('maximum herd home range'). The log-normal distribution was parameterized with the mean and coefficient of variation (CV) of the observed scale, where the standard deviation of the log-normal variable (SD_{ln}) is given by:

$$SD_{ln} = \sqrt{\ln(1 + CV^2)} \qquad \text{eqn 4}$$

and the mean of the log-normal variable (mean$_{ln}$) is given by:

$$\text{mean}_{ln} = \ln(\text{mean}) - 0.5\ln(1 + CV^2) \qquad \text{eqn 5}$$

The log-normal distribution is commonly applied to describe multiplicative processes when mean values are low, variances large, and values cannot be negative, such as wildlife abundance (Matthiopoulos 2011). Impala monitoring does not provide counts of juveniles so population sizes were assumed to refer to adults only.

WILDLIFE MONITORING

The observation procedure was modelled according to descriptions of monitoring in the Serengeti by Sinclair (1973), Norton-Griffiths (1973, 1978), Sinclair & Norton-Griffiths (1982), and Hilborn & Sinclair (2010). Therefore, we simulated monitoring of wildebeest in the Serengeti plains during the wet season when animals are aggregated in large numbers in a relatively small area. This timing increases the likelihood of good visibility and leads to a clear separation between migratory and resident wildebeest (only migratory animals are surveyed in this monitoring design). Migratory wildebeest are monitored through aerial point sampling (APS; Norton-Griffiths 1978, 1988), which was first developed to characterize the land use of a region by interpreting a sample of aerial photographs. We simulated monitoring of impala using Systematic Reconnaissance Flight (SRF) surveys, in which abundance data are collected within each subunit along several flight lines (i.e. transects; Norton-Griffiths 1978).

SAMPLING

The sampling steps simulated the process of taking aerial photographs of wildebeest or conducting direct impala counts through selection of grid cells. We simulated a wide range of levels of sampling effort (measured by distances between transects and spacing between photographs) to investigate the effects of sampling error on survey accuracy and precision. Detailed information about each specific step in this model component is provided in the online supporting information (Appendix S1, Supporting information).

Fig. 2. Estimated probability of juvenile presence according to total number of animals per photograph. Original data on the presence of juveniles are superimposed as grey circles, with diameter proportional to the total number of animals. The trend line represents effect taken from model outputs (GLM with binomial errors, $N = 343$ photographs) and the dashed lines indicate 95% confidence intervals.

OBSERVATIONAL PROCEDURES: FLIGHT AND OBSERVER EFFECTS

These steps simulated flight and counting characteristics, namely variation in flight altitude and speed, and spatially autocorrelated

miscounting of animals from photographs or direct counts because of bad weather conditions, habitat or fatigue. For impala, observer effects also included herd and individual detectability and distance. This component allowed us to investigate the effects of observation error on survey accuracy and precision. Detailed information is provided in the online supporting information (Appendix S2, Supporting information).

A summary of the modelled variables and range of values explored for each species is presented in Table 1.

ABUNDANCE ESTIMATION

The simulated survey data were used to estimate wildlife abundance, following procedures currently adopted in the study area.

Wildebeest

Simulated aerial photographs were treated as simple random samples from which juvenile and adult wildebeest were counted (Hilborn & Sinclair 2010). The estimated density per photograph is the number of animals in a certain photograph divided by the photograph area. The final estimate of the wildebeest population size is the area included in the survey (90×35 km^2) multiplied by the average density:

$$N = DA \qquad \text{eqn 6}$$

Impala

Data from a simulated SRF were converted to estimates of animal density by dividing the total number of animals seen by both observers by the length of the subunit multiplied by their summed strip widths (Norton-Griffiths 1978). Subunits were then combined within each transect and population estimates were calculated using transects as units of random sampling (Campbell & Borner 1995).

ANALYSIS OF SOURCES OF OBSERVATION UNCERTAINTY

Analysis involved varying all model parameters simultaneously within the range considered (Table 1) and testing their effects on survey accuracy and precision. Survey precision was measured by the CV and the normalized variance (CV is squared to the coefficient of variance, CV^2, which represents the total observed variance) but only CV results are presented (CV^2 results in Supporting information). The CV was based on the simulated survey data, rather than statistically derived from each survey estimate, and expressed as:

$$CV = \frac{SD}{\bar{x}} \qquad \text{eqn 7}$$

where SD is the sample standard deviation of the population estimates from 50 simulations and $\bar{x}$ the mean estimate of population size. Accuracy was defined as the percentage discrepancy between the mean estimated population size and the known population sizes for juveniles only or all age classes together (for impala, only adult counts were conducted).

One thousand sets of parameter values were generated independently from uniform distributions for each species, and 50 simulations were carried out for each parameter set, from which mean values were obtained. All explanatory and dependent variables were scaled to have a standard deviation of unity, resulting in unit-less measures that can be used to infer the relative importance of parameters. Generalized linear models with Gamma (loglink) and Gaussian error distributions were fitted to the simulation results to evaluate the sensitivity of survey precision and accuracy to parameters, respectively. A generalized linear model with quasi-binomial error distribution (to account for overdispersion) and a logit link was fitted for juvenile wildebeest. Relevant two-way interactions were also considered. The linearity of the relationship between the parameters and the dependent variables and model residuals was examined graphically.

We also explored under which conditions (population characteristics and observation error) SRFs were adequate for impala monitoring. We assumed that at least one herd or five animals would have to be seen in order for the method to be considered adequate and used a generalized linear model with binomial error distribution to evaluate the effects of potential drivers on survey adequacy, treated as a binary variable.

Finally, to illustrate the potential effects of different budget allocations on survey precision and accuracy, we ran 50 replicates for 1000 parameter sets under high- and low-budget scenarios. High- or low-budget scenarios assume parameters at their best or worst values, respectively (values presented in Table S1, Supporting information). For example, the low-budget scenario assumes that only a few transects are conducted and that there is high counting variability (perhaps due to inexperienced or untrained observers). We obtained current unitary costs from itemized monitoring expenses in the study area (J.G. C. Hopcraft, unpublished data) and then multiplied them by the simulated parameter values to estimate approximate budget costs for both scenarios.

Results

EFFECTS OF SURVEY CHARACTERISTICS ON PRECISION AND ACCURACY

Wildebeest

As wildebeest became more aggregated (i.e. lower k values) and more spatially autocorrelated (i.e. higher similarity between nearby cells), the surveys became less precise (higher CV). Higher sampling effort (i.e. smaller distance between transects or spacing between photographs) increased precision in the wildebeest surveys but this effect was significantly weaker when spatial autocorrelation increased (Table 2).

The comparison between the population estimates from the surveys and known population sizes suggested that accuracy was lower for higher population sizes, when juveniles constituted a higher proportion of the total population, and for lower levels of juvenile detectability when counting from photographs (Table 2). Accuracy of juvenile estimates was mostly affected by population size, juvenile detectability and aggregation.

Table 1. Description of variables and range of values explored for monitoring of (A) wildebeest and (B) impala. The subscripts 'wild' and 'imp' refer to parameters regarding wildebeest and impala, respectively

Parameters	Notation	Range	Sources
(A) Wildebeest			
Population characteristics			
Population size	N_{wild}	200 000–2 000 000	Hilborn & Sinclair (2010)
Proportion of juveniles (%)	α	5–35	Estimated
Aggregation	k	0·01–2	Assumed
Spatial autocorrelation range	phi	0·1–0·5	Assumed
Spatial threshold variance (sill)	σ_s^2	Fixed (1)	Assumed
Sampling characteristics			
Distance between transects (km)	γ_{wild}	0·5–24	Hilborn & Sinclair (2010)
Time between photographs (s)	δ	1–120	Hilborn & Sinclair (2010)
Flight characteristics			
Mean flight altitude (feet)	ε_{wild}	Fixed (1200)	Hilborn & Sinclair 2010;
CV error altitude	ζ_{wild}	0–0·2	Estimated
Mean flight speed (km s^{-1})	θ_{wild}	Fixed (0·06)	Hilborn & Sinclair 2010;
CV error speed	ι_{wild}	0–0·3	Assumed
Observer effects			
Minimum error counting juveniles (%)	Λ	0–0·2	Sinclair (1973)
Juvenile detectability (number of animals in a photograph for which 50% juveniles are likely to be missed)	$\ddot{\imath}$	20–50	Assumed
CV error counting adults	$\ddot{v}$	0–0·5	Assumed
Counting error autocorrelation range	phi$_{wild}$	0–1	Assumed
(B) Impala			
Population characteristics			
Population size	N_{imp}	1000–15 000	Grumeti Fund (2010)
Median herd size	ξ	5–50	Jarman & Jarman (1973)
CV herd size	o	0–0·5	Stein & Georgiadis (2008)
Maximum herd home range (km^2)	π	0·5–3	Jarman & Sinclair (1979)
Individual space (km^2)	ς	0·05–0·2	Jarman & Sinclair (1979)
Sampling characteristics			
Distance between transects (km)	γ_{imp}	0·5–7	TAWIRI (2010)
Flight characteristics			
Mean flight altitude (feet)	ε_{imp}	Fixed (300)	TAWIRI (2010)
CV error altitude	ζ_{imp}	0–0·2	Assumed
Mean flight speed (km s^{-1})	θ_{imp}	Fixed (0·06)	TAWIRI (2010)
Observer effects			
Minimum herd detectability (%)	σ	0·05–0·5	Assumed
Herd size nondetectability (herd size for which there is a 50% chance of missing it)	τ	10–50	Assumed
Individual detectability at distance 0 (%)	υ	0·7–0·99	Assumed
Detectability by distance (distance for which there is a 50% chance of seeing animals; km)	φ	0·125–0·250	Assumed
Maximum individual detectability (%)	χ	0·7–0·99	Assumed
Herd size estimability (number of animals in a herd for which 50% are likely to be missed)	ψ	10–50	Assumed
CV counting error	ω	0–0·5	Assumed
CV counting error autocorrelation range	phi$_{imp}$	0–1	Assumed

CV, coefficient of variation.

Impala

For impala monitoring through SRFs, the surveys became less precise as distance between transects increased (i.e. lower sampling effort), for lower population sizes, higher mean herd sizes and lower herd size estimability. Accuracy in SRFs decreased when detectability at minimum distance and herd size estimability decreased and mean herd size and herd size nondetectability increased (Table 3).

The likelihood of detecting at least one herd or five animals using SRFs decreased for lower population sizes, lower sampling effort (measured as distance between transects), higher mean herd size, lower maximum individual detectability and lower herd size estimability (Table 3).

BUDGETARY SCENARIOS

High budgets produced more precise estimates for wildebeest and impala (Fig. 3b, d) although both monitoring

Table 2. Results of a sensitivity analysis in which generalized linear models were fitted to precision (coefficient of variation) and inaccuracy (percentage discrepancy between the mean estimated population size and the known population size) for wildebeest monitoring

Parameter	Relative importance (standardized regression coefficients; β)		
	Coefficient of variation (CV)	Inaccuracy	Inaccuracy (juveniles only)
Population size	−0·03	**0·40*****	**0·70*****
Proportion of juveniles	−0·02	**0·71*****	0·05***
Aggregation (k)	**−0·17*****	−0·08***	**−0·12*****
Spatial autocorrelation	**0·35*****	−0·04*	0·01
Distance between transects (km)	**0·13*****	−0·01	−0·09**
Time between photographs (s)	**0·14*****	−0·01	−0·08**
CV error altitude	0·02	0·04	−0·02
CV error speed	−0·01	0·01	−0·01
Minimum error counting juveniles (%)	0·02	0·05***	−0·01
Juvenile detectability	−0·02	**−0·16*****	**−0·29*****
CV of error counting adults	0·03**	0·01	0·02
Spatially autocorrelated errors	0·01	−0·03	−0·01
Spatially autocorrelated error*CV of error counting adults	−0·02	0·01	0·01
Spatially autocorrelated error*Juvenile detectability	−0·03	0·02	−0·01
Spatially autocorrelated error*minimum error counting juveniles	0·01	−0·01	0·01
Spatial autocorrelation*distance between photographs	−0·07**	0·02	−0·01
Spatial autocorrelation*distance between transects	−0·09**	0·01	−0·01
Aggregation*Spatial autocorrelation	0·05	−0·02	−0·01

All dependent and explanatory variables were scaled to have a standard deviation of unity for comparative purposes. The table shows the coefficients of all parameters and interactions from the full model. All β > 0·10 are given in bold. Significance is coded as ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

Table 3. Results of a sensitivity analysis in which generalized linear models were fitted to precision (coefficient of variation), inaccuracy (percentage discrepancy between the mean estimated population size and the known population size) and survey adequacy (able to detect at least one herd or five animals) for impala monitoring

Parameter	Relative importance (standardized regression coefficients; β)		
	Coefficient of variation (CV)	Inaccuracy	Adequacy
Population size	**−0·32*****	0·05*	**0·36*****
Mean herd size	**0·15*****	**0·25*****	**−0·31*****
CV herd size	0·02	0·10***	0·01
Maximum herd home range (km²)	−0·04	0·03	0·04
Individual space (km²)	−0·03	−0·01	−0·01
Distance between transects (km)	**0·45*****	0·03	**−0·53*****
CV error altitude	−0·03	−0·02	−0·04
Minimum herd detectability (%)	−0·05	−0·07***	0·04
Herd size nondetectability	0·08**	**0·39*****	−0·03
Detectability at distance 0 (%)	−0·02	**−0·20*****	0·07
Detectability by distance	−0·03	−0·04*	0·02
Maximum individual detectability (%)	−0·01	−0·04	**−0·18*****
Herd size estimability	**−0·11*****	**−0·62*****	**0·54*****
CV counting error	0·02	−0·01	0·01
Spatially autocorrelated errors	0·01	0·01	−0·01
Spatially autocorrelated errors*CV counting error	0·01	0·01	−0·01

All dependent and explanatory variables were scaled to have a standard deviation of unity for comparative purposes. The table shows the coefficients of all parameters and interactions from the full model. All β > 0·10 are given in bold. Significance is coded as ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

techniques were likely to underestimate wildlife abundance (Fig. 3a, c). Mean underestimation for wildebeest monitoring from APS was around 15% and, although low- and high-budget scenarios produced similar values of mean underestimation, survey accuracy was much more variable for low budgets, producing estimates from 60% below the known population size up to 30% above (Fig. 3a).

For impala monitoring, the SRFs produced estimates generally 80% below the true value. Higher budget sizes produced a mean underestimation of around 65% and
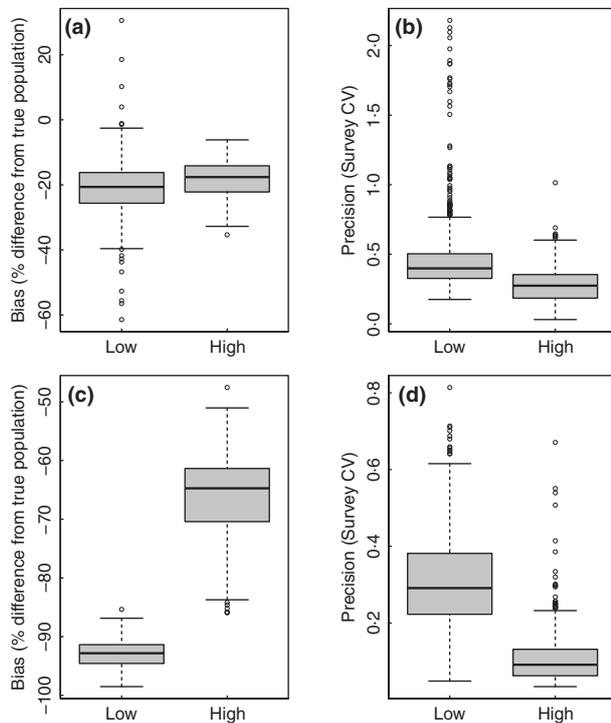
Fig. 3. The potential effects of different budget allocations (low- or high-budget scenarios) on (a) survey accuracy for wildebeest monitoring; (b) survey precision for wildebeest monitoring; (c) survey accuracy for impala monitoring; and (d) survey precision for impala monitoring. High- or low-budget scenarios assume parameters at their best or worst values, respectively (see Table S1, Supporting information). For example, the low-budget scenario assumes conducting only a few transects and high counting variability.

lower budget sizes produced a mean underestimation of around 90%, ranging from 80% to 100%. Higher budgets produced more variable bias, unlike for the wildebeest; this is due to variability being constrained by reaching zero herds detected in the low-budget scenario.

For wildebeest monitoring, the simulated high- and low-budget scenarios were estimated to cost *c.* US$6410 and US$3780, respectively. For impala, the high budget would require around US$8250 and the lower *c.* US$4250. The higher-budget scenario considered for wildebeest monitoring, which provides twice more precise estimates, would thus cost *c.* 70% more to implement. The higher budget for impala monitoring, which increases accuracy by 30%, would increase costs roughly by 95%.

Discussion

In this study, we have considered the multiple sources and effects of uncertainty in monitoring data obtained through wildlife surveys, focusing on the interactions between observation error and the spatial distribution of wildlife populations. Our results suggest that, under the simulated conditions, the relative importance of each process affecting precision and accuracy varies according to the survey technique and biological characteristics of the

species. While survey precision was mainly affected by population characteristics and sampling effort, the accuracy of the survey was greatly affected by observer effects, such as juvenile and herd detectability. The adequacy of SRFs, that is, whether these surveys led to a minimum number of sightings, was mainly affected by population size, mean herd size, herd size estimability, maximum individual detectability and sampling effort. Our results also illustrate how budget size affects survey precision and accuracy, particularly for SRFs.

We extend previous work on causes of survey bias and imprecision (e.g. Norton-Griffiths 1978; Norton-Griffiths & McConville 2007) by developing a 'virtual ecologist' framework (Zurell *et al.* 2010) within which to carry out simulated tests of different monitoring strategies for different types of species. Elphick (2008) highlights the need for improved understanding of the effects of multiple sources of uncertainty on survey bias and precision, particularly errors due to observation uncertainty and its interaction with biological characteristics. However, compared to other aspects of monitoring such as sampling design, observation uncertainty is still the 'Cinderella' of monitoring, with little attention to the multiple potential sources of error involved. By decomposing observation uncertainty into components, which may vary in magnitude and direction, we can make practical recommendations to managers concerning the priority issues that require attention. This would allow them to improve precision or accuracy of their counts, depending on the biology of the species concerned and budgetary constraints (Table 4).

The spatial distribution of a species is a major driver of variation in survey precision and accuracy (Table 4). Our findings chime with those of, for example, Blanchard, Maxwell & Jennings (2008) and Borkowski, Palmer & Borowski (2011), who also show the importance of aggregation (due to biological/social characteristics) and spatial autocorrelation (due to environmental/spatial characteristics) in determining survey precision. Counterbalanced variation due to changes in sampling effort, aggregation and spatial autocorrelation (for more aggregated species), and population size and mean herd size (for less aggregated species) suggests that sampling effort should be defined according to the spatial distribution to account for differences in precision. For monitoring highly aggregated species, such as wildebeest, we recommend that particular attention should be given to survey precision and that sampling effort should be defined according to previous estimates of aggregation in the monitored population. For example, in the Serengeti, sampling effort varies between years according to rough visual estimations of aggregation. This assessment could be formally considered in the monitoring protocol. The survey precision is most sensitive to spatial autocorrelation, which should be explicitly considered in abundance estimation procedures (e.g. confidence levels adjusted for 'effective sample size' lower than actual sample size).

Table 4. Summary of the main issues considered in this study and our main recommendations for different types of species according to their spatial distribution, listed in priority order

	Type of species according to spatial distribution	
	Highly aggregated (e.g. wildebeest)	Random or slightly aggregated (e.g. impala)
Aerial survey technique analysed	Aerial Point Sampling	Systematic Reconnaissance Flights
Main issues considered	Sampling effort	Sampling effort
	Flight characteristics (variation in altitude and speed)	Flight characteristics (variation in altitude)
	Spatial distribution (aggregation and spatial autocorrelation)	Spatial distribution (herd size and home range)
	Population size and structure (proportion of juveniles)	Population size
	Observer effects (juvenile detectability and counting error of adult animals)	Observer effects (counting error, herd detectability according to size, individual detectability within herd and distance effects)
Prioritized recommendations	Focus on survey precision	Focus on survey bias
	Obtain preliminary estimates of aggregation and spatial autocorrelation, and define sampling effort accordingly	Maximize, and obtain estimates of, herd size estimability
	Minimize, and obtain estimates of, counting errors of juvenile animals or obtain juvenile estimates from ground transects	Maximize, and obtain estimates of, herd detectability
		Apply bias correction factor according to mean herd size

Similarly to other studies comparing estimates obtained through aerial surveys to known or presumed accurate population sizes (Goddard 1967; Jachmann 2002), our simulated surveys produced underestimates of considerable magnitude. Survey accuracy was greatly affected by multiple observer effects, particularly juvenile detectability when counting from photographs, and herd size estimability and detectability when conducting direct counts during transects. Although the effects of distance and counting variability have been often mentioned as sources of inaccuracy (Buckland 2001), our results show that these commonly discussed types of observer error were comparatively less important in driving survey accuracy for these species in the range of conditions that occur in the Serengeti. This demonstrates the need for error minimization priority-setting based on comparative analyses. For example, McConville *et al.* (2009) explore the effect of herd detectability on accuracy, but we show that aerial survey accuracy is very much affected by detectability of individual animals within a herd. For random or slightly aggregated species monitored through aerial surveys, such as impala, we recommend that minimising potential bias should be a major consideration. As accuracy is most sensitive to observer effects, monitors should be provided with appropriate training and their reliability evaluated before the actual survey to calibrate the final abundance estimates. For example, observers' estimates could be compared with photographs of herds, obtaining correction factors. Other studies have shown that ground counts can provide more accurate estimates than aerial surveys, which are greatly affected by wildlife visibility for this type of species, but are generally more time-consuming and expensive, particularly for large survey areas (Jachmann 2002). When feasible, ground counts, or other

better-performing techniques, should be conducted instead of or in addition to aerial surveys.

We also highlight the importance of considering which demographic group is subject to biases. In the case of wildebeest, juvenile detectability was a key driver of survey accuracy, while the effect of miscounting adults was negligible. Variation in juvenile survival can be used to make inferences about population trends (Gaillard, Festa-Bianchet & Yoccoz 1998), which further illustrates the importance of correctly counting juveniles. For highly aggregated populations, juvenile abundance could be obtained from other sources, such as ground transects, to avoid reducing accuracy of total population estimates. In other species, there may be different population components for which accurate and precise abundance estimates are crucial to management. For example, Katzner, Milner-Gulland & Bragin (2007) demonstrated the importance of collecting data on adult survival of Imperial eagles *Aquila heliaca* instead of territory occupancy to detect population trends.

This study took a static, spatially explicit approach to analysing monitoring uncertainties in the range of conditions that occur in the Serengeti, but there are also issues related to changes over time. For example, observer performance may improve or herd aggregation coefficients may change (cf McConville *et al.* 2009). Chee & Wintle (2010) have developed a dynamic cull control rule for overabundant wildlife where iterative culling can be used to update population parameters through Bayesian methods. Similarly, a dynamic monitoring strategy could update according to knowledge gained from the observation process.

Monitoring efficiency is of the utmost importance for conservation especially in the context of limited budgets

and other priorities (Danielsen *et al.* 2003; Bottrill *et al.* 2008). Relating data quality to budgetary constraints for different survey techniques and prioritizing approaches to error minimization are thus essential to investigate trade-offs and make informed decisions under uncertainty (Caughlan & Oakley 2001; Gaidet-Drapier *et al.* 2006) but these are rarely considered.

Monitoring and management decisions should be incorporated into conceptual and methodological frameworks, which explicitly consider uncertainty, such as the management strategy evaluation (MSE; Bunnefeld, Hoshino & Milner-Gulland 2011) and adaptive management (AM; Keith *et al.* 2011). MSE uses monitoring data to estimate trends and population size and then simulates decisions taking the degree of observation uncertainty into account, while AM implements strategies that incorporate uncertainty by testing multiple plausible hypotheses. Using a 'virtual ecologist' approach (Zurell *et al.* 2010), we provided insights into how to improve monitoring data and implement informed management actions that take monitoring uncertainty into consideration. This approach could easily be integrated into an MSE or AM framework. Explicit analyses of multiple types and sources of uncertainty are required, ensuring that conservation trade-offs are evaluated in a comprehensive, robust and transparent manner (Chee & Wintle 2010).

Acknowledgements

References

Blanchard, J., Maxwell, D.L. & Jennings, S. (2008) Power of monitoring surveys to detect abundance trends in depleted populations: the effects of density-dependent habitat use, patchiness, and climate change. *ICES Journal of Marine Science*, **65**, 111–120.

Borkowski, J., Palmer, S.C.F. & Borowski, Z. (2011) Drive counts as a method of estimating ungulate density in forests: mission impossible? *Acta Theriologica*, **56**, 239–253.

Bottrill, M.C., Joseph, L.N., Carwardine, J., Bode, M., Cook, C., Game, E.T., Grantham, H., Kark, S., Linke, S., McDonald-Madden, E., Pressey, R.L., Walker, S., Wilson, K.A. & Possingham, H.P. (2008) Is conservation triage just smart decision making? *Trends in Ecology & Evolution*, **23**, 649–654.

Buckland, S.T. (2001) *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford, UK.

Bunnefeld, N., Hoshino, E. & Milner-Gulland, E.J. (2011) Management strategy evaluation: a powerful tool for conservation? *Trends in Ecology & Evolution*, **26**, 441–447.

Campbell, K. & Borner, M. (1995) Population trends and distribution of Serengeti herbivores: implications for management. *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds K.L.I. Campbell & M. Borner), pp. 117–145. University of Chicago Press, Chicago.

Caughlan, L. & Oakley, K.L. (2001) Cost considerations for long-term ecological monitoring. *Ecological Indicators*, **1**, 123–134.

Chee, Y.E. & Wintle, B.A. (2010) Linking modelling, monitoring and management: an integrated approach to controlling overabundant wildlife. *Journal of Applied Ecology*, **47**, 1169–1178.

Danielsen, F., Mendoza, M.M., Alviola, P.A., Balete, D.S., Enghoff, M., Poulsen, M.K. & Jensen, A.E. (2003) Biodiversity monitoring in developing countries: what are we trying to achieve? *Oryx*, **37**, 407–409.

Diggle, P.J., Tawn, J.A. & Moyeed, R.A. (1998) Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **47**, 299–350.

Elphick, C.S. (2008) How you count counts: the importance of methods research in applied ecology. *Journal of Applied Ecology*, **45**, 1313–1320.

Estes, R. (1992) *The Behavior Guide to African Mammals: Including Hoofed Mammals, Carnivores, Primates*. University of California Press, Berkeley, USA.

Field, S.A., Tyre, A.J. & Possingham, H.P. (2005) Optimizing allocation of monitoring effort under economic and observational constraints. *The Journal of Wildlife Management*, **69**, 473–482.

Gaidet-Drapier, N., Fritz, H., Bourgarel, M., Renaud, P.-C., Poilecot, P., Chardonnet, P., Coid, C., Poulet, D. & Bel, S. (2006) Cost and efficiency of large mammal census techniques: comparison of methods for a participatory approach in a communal area, Zimbabwe. *Biodiversity and Conservation*, **15**, 735–754.

Gaillard, J.-M., Festa-Bianchet, M. & Yoccoz, N.G. (1998) Population dynamics of large herbivores: variable recruitment with constant adult survival. *Trends in Ecology & Evolution*, **13**, 58–63.

Goddard, J. (1967) The validity of censusing black rhinoceros populations from the air. *East African Wildlife Journal*, **5**, 18–23.

Grumeti Fund. (2010) *Singita Grumeti Community & Wildlife Conservation Fund: 2002–2010*. Grumeti Fund, Unpublished report.

Harwood, J. & Stokes, K. (2003) Coping with uncertainty in ecological advice: lessons from fisheries. *Trends in Ecology & Evolution*, **18**, 617–622.

Hilborn, R. & Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, USA.

Hilborn, R. & Sinclair, A.R.E. (2010) *Summary of Wildebeest 2009 Count Evaluation*. Frankfurt Zoological Society, Unpublished report.

Jachmann, H. (2002) Comparison of aerial counts with ground counts for large African herbivores. *Journal of Applied Ecology*, **39**, 841–852.

Jarman, M.V. & Jarman, P.J. (1973) Daily activity of impala. *African Journal of Ecology*, **11**, 75–92.

Jarman, P.J. & Sinclair, A.R.E. (1979) Feeding strategy and the pattern of resource partitioning in ungulates. *Serengeti: Dynamics of an Ecosystem* (eds A.R.E. Sinclair & M. Norton-Griffiths), pp. 130–163. University of Chicago Press, Chicago.

Katzner, T., Milner-Gulland, E.J. & Bragin, E. (2007) Using modeling to improve monitoring of structured populations: are we collecting the right data? *Conservation Biology*, **21**, 241–252.

Keith, D.A., Martin, T.G., McDonald-Madden, E. & Walters, C. (2011) Uncertainty and adaptive management for biodiversity conservation. *Biological Conservation*, **144**, 1175–1178.

Legg, C.J. & Nagy, L. (2006) Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management*, **78**, 194–199.

Matthiopoulos, J. (2011) *How to Be a Quantitative Ecologist: The "a to R" of Green Mathematics and Statistics*. John Wiley and Sons, Chichester, UK.

McConville, A., Grachev, I., Keane, A., Coulson, T., Bekenov, A.B. & Milner-Gulland, E.J. (2009) Reconstructing the observation process to correct for changing detection probability of a critically endangered species. *Endangered Species Research*, **6**, 231–237.

McDonald-Madden, E., Baxter, P.W.J., Fuller, R.A., Martin, T.G., Game, E.T., Montambault, J. & Possingham, H.P. (2010) Monitoring does not always count. *Trends in Ecology & Evolution*, **25**, 547–550.

Milner-Gulland, E.J., Coulson, T. & Clutton-Brock, T.H. (2004) Sex differences and data quality as determinants of income from hunting red deer *Cervus elaphus*. *Wildlife Biology*, **10**, 187–201.

Nichols, J.D. & Williams, B.K. (2006) Monitoring for conservation. *Trends in Ecology & Evolution*, **21**, 668–673.

Norton-Griffiths, M. (1973) Counting the Serengeti migratory wildebeest using two-stage sampling. *African Journal of Ecology*, **11**, 135–149.

Norton-Griffiths, M. (1978) *Counting Animals*. African Wildlife Leadership Foundation, Nairobi, Kenya.

Norton-Griffiths, M. (1988) Aerial point sampling for land use surveys. *Journal of Biogeography*, **15**, 149–156.

Norton-Griffiths, M. & McConville, A. (2007) *Technical Evaluation of the Aerial Survey Programme for Saiga Antelope (Saiga Tatarica) With*

Recommendations for Improvements. Final Report. ZGF-Project No. 1293/02 Saiga Schutz Kasachstan/Evaluation Der Tierzahlungen. Frankfurt Zoological Society, Frankfurt.

Perry, J.N. & Taylor, L.R. (1986) Stability of real interacting populations in space and time: implications, alternatives and the negative binomial kc. *Journal of Animal Ecology*, **55**, 1053–1068.

Pieters, E.P., Gates, C.E., Matis, J.H. & Sterling, W.L. (1977) Small sample comparison of different estimators of negative binomial parameters. *Biometrics*, **33**, 718–723.

Reynolds, J.H., Thompson, W.L. & Russell, B. (2011) Planning for success: identifying effective and efficient survey designs for monitoring. *Biological Conservation*, **144**, 1278–1284.

Ribeiro, P.J. Jr & Diggle, P.J. (2001) geoR: a package for geostatistical analysis. *R-News*, **1**, 15–18.

Sethi, G., Costello, C., Fisher, A., Hanemann, M. & Karp, L. (2005) Fishery management under multiple uncertainty. *Journal of Environmental Economics and Management*, **50**, 300–318.

Shea, K. & Mangel, M. (2001) Detection of population trends in threatened Coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 375–385.

Sims, M., Wanless, S., Harris, M.P., Mitchell, P.I. & Elston, D.A. (2006) Evaluating the power of monitoring plot designs for detecting long-term trends in the numbers of common guillemots. *Journal of Applied Ecology*, **43**, 537–546.

Sinclair, A.R.E. (1973) Population increases of buffalo and wildebeest in the Serengeti. *African Journal of Ecology*, **11**, 93–107.

Sinclair, A.R.E. & Norton-Griffiths, M. (1982) Does competition or facilitation regulate migrant ungulate populations in the Serengeti? A test of hypotheses. *Oecologia*, **53**, 364–369.

Sinclair, A.R.E., Mduma, S., Hopcraft, J.G.C., Fryxell, J.M., Hilborn, R. & Thirgood, S.J. (2007) Long-term ecosystem dynamics in the Serengeti: lessons for conservation. *Conservation Biology*, **21**, 580–590.

Stein, A. & Georgiadis, N. (2008) Spatial statistics to quantify patterns of herd dispersion in a savanna herbivore community. *Resource Ecology: Spatial and Temporal Dynamics of Foraging* (eds H.H.T. Prins & F.V. Langevelde), pp. 33–52. Springer, Dordrecht, The Netherlands.

Stem, C., Margoluis, R., Salafsky, N. & Brown, M. (2005) Monitoring and evaluation in conservation: a review of trends and approaches. *Conservation Biology*, **19**, 295–309.

TAWIRI. (2010) *Aerial Census in the Serengeti Ecosystem – Wet Season 2010*. TAWIRI, Arusha, Tanzania.

Thirgood, S.J., Mosser, A., Tham, S., Hopcraft, J.G.C., Mwangomo, E., Mlengeya, T., Kilewo, M., Fryxell, J.M., Sinclair, A.R.E. & Borner, M. (2004) Can parks protect migratory ungulates? The case of the Serengeti wildebeest. *Animal Conservation*, **7**, 113–120.

Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, **16**, 446–453.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1. Map of the study area.

Appendix S1. Description of the sampling steps in the wildlife monitoring model.

Appendix S2. Description of the observational procedures in the wildlife monitoring model.

Appendix S3. R code used for simulations.

Table S1. Description of modelled variables and range of values explored for low and high-budget monitoring scenarios.

Table S2. Results from generalized linear models fitted to precision and inaccuracy for wildebeest monitoring.

Table S3. Results from generalized linear models fitted to precision, inaccuracy and survey adequacy for impala monitoring.