Ecol Res (2009) 24: 345–353
DOI 10.1007/s11284-008-0511-3

ORIGINAL ARTICLE

Thomas C. Wanger · Iris Motzke · Samuel C. Furrer
Barry W. Brook · Bernd Gruber

How to monitor elusive lizards: comparison of capture–recapture methods on giant day geckos (*Gekkonidae, Phelsuma madagascariensis grandis*) in the Masoala rainforest exhibit, Zurich Zoo

Received: 27 September 2007 / Accepted: 5 May 2008 /

Abstract Rapid and reliable estimation of population size is needed for the efficient monitoring of animal populations of conservation concern. Unfortunately, technical advances in this area have not been paralleled in uptake in conservation, which may be due to difficulties in implementation or the lack of general guidelines for application. Here we tested five different methods used to estimate population size [capture–mark–recapture (CMR), finite-mixture models, model averaging of finite-mixture models, accumulation curve methods (ACM), and the line transect method (LT)] using extensive capture–recapture data of the giant day gecko (*Gekkonidae*, *Phelsuma madagascariensis grandis*, Gray 1870) at the Masoala rainforest exhibit, Zurich Zoo. When the complete data were analyzed [30 sessions (and 27 sessions for the LT)], all methods except the LT produced similar estimates of population size. The simple ACM gave a small coefficient of variation (CV), but did not cover the most likely value of population size at moderate sampling effort. Nevertheless, the ACM was the only method that showed a reasonable convergence when subsets of data were used. CMR and Pledger models included the reference value in their confidence intervals (CI) after 25 and 30 sessions, respectively. Although model averaging did slightly improve the estimate, the CV was still high for the full dataset. Our method of using subsets of data to test the robustness of estimates is simple to apply and could be adopted more widely in such analyzes to evaluate sensitivity to method of evaluation. In conclusion, simple accumulation methods showed similar efficiency to more complex statistical models, and are likely to be sufficiently precise for most conservation monitoring purposes.

Keywords Capture–mark–recapture · Finite-mixture models · Model averaging · Accumulation curve method · Line transect method · Method comparison

Introduction

Over the last 40 years capture–mark–recapture (CMR) theory and accumulation curve methods (ACM), also called rarefaction, have been a field of intense research, with the goal of improving estimates of density and abundance in biological populations (Kohn et al. 1999; Schwarz and Seber 1999; Buckland et al. 2001; Williams et al. 2002; Lukacs and Burnham 2005; Petit and Valiere 2006). These methods, based on numerous theoretical assumptions, have been used for a wide range of animal taxa to estimate population size and demographic parameters (e.g., Palsbøll et al. 1997; Schaub et al. 2001; Eggert et al. 2003; Silver et al. 2004). Their implementation is now supported by freely available computer

Electronic supplementary material The online version of this article (doi:10.1007/s11284-008-0511-3) contains supplementary material, which is available to authorized users.

T. C. Wanger · I. Motzke
Department of Zoology, University of Tübingen,
Tübingen, Germany

S. C. Furrer
Zoo Zurich, Zürichbergstrasse 221,
8044 Zurich, Switzerland

T. C. Wanger (✉) · B. W. Brook
Research Institute for Climate Change and Sustainability,
School of Earth and Environmental Sciences,
University of Adelaide, Adelaide, SA 5005, Australia
E-mail: thomas.wanger@adelaide.edu.au
Tel.: +61-8-83035254

B. Gruber
Department for Computational Landscape Ecology,
Centre for Environmental Research, UFZ, Leipzig, Germany

programs such as MARK and CAPTURE (White and Burnham 1999). In herpetology, CMR methods are often used (e.g., Wood et al. 1998; Freilich et al. 2000; Tyrell et al. 2000; Savarie et al. 2001; Gruber and Henle 2004) but validation of population estimates has rarely been attempted.

As research in this field is dynamic, new methods are regularly developed and old ones improved (e.g., Coull and Agresti 1999; Chao et al. 2000; Yip et al. 2000; Chao 2001; Pledger 2005). As such, ecologists are faced with a "moving target" problem of choosing the most adequate method for their purpose. For effective monitoring of animal populations, especially those of conservation concern that are usually difficult to monitor, the most efficient sampling regime must be identified (e.g., number of re-sampling sessions). Yet because there is no standard way of estimating population size from capture–recapture data, validation exercises are necessary to understand how a statistical estimator compares to the true population size. One way to validate an estimator is against a proven reference value (i.e., a direct count of total population size). An individual census, however, is extremely difficult to achieve under natural conditions and, thus, rigorous tests of different methods with field data are scarce.

We tested five different methods, CMR, Pledger's finite-mixture models, model averaging based on the Pledger models, ACM, and LT on a realistic and well-studied model species, the arboreal giant day gecko (*Gekkonidae*, *Phelsuma madagascariensis grandis*, Gray 1870). In an intensive capture–recapture study, we conducted 30 recapture sessions (and 27 sessions for LT) in the Masoala rainforest exhibit, Zurich Zoo, and used the individual color patterns of this gecko species for photo recognition. Due to the high number of sampling sessions, this study is based on an unusually good dataset. Moreover, it is a captive population and, therefore, we can be sure that the closed population assumption is valid. However, the dataset is comparable to natural conditions because of the dimensions of the Masoala rainforest exhibit; it is currently the second largest tropical exhibit in the world. Additional variation in the dataset results from time-dependent variation in recapture rates (obvious dependence of gecko activity on daily weather) and the photographic capture method (it is harder to spot and photograph juvenile compared to adult geckos).

The aim of this study was to identify the most efficient method of estimating the size of this gecko population, as an exemplar for field monitoring of arboreal lizards. For the method assessment, we firstly calculated the coefficient of variation (CV) for the full data set, to evaluate the achieved precision. We then compared the performance of different methods with subsets of the data to test the theoretical expectation that differences between methods in population size estimation should become smaller when sample sizes increase.

Materials and methods

Study area and species

The Masoala rainforest exhibit of Zurich Zoo covers an area of almost 11,000 m^2 and has a height of 35 m. More than 45 animal species of different taxa and over 35,000 individual plants have been kept here under semi-natural conditions since 2003 (Furrer et al. 2006).

The giant day gecko (*Phelsuma madagascariensis grandis*) is a tropical reptile naturally occurring in the northern part of Madagascar. As this species shows only slight sexual dimorphism (males are longer and have a larger head-width than females; Hallmann et al. 1997), it was impossible to sex adults without catching them. We classified juveniles as 1/3 total length of an adult (i.e., 30 cm; Henkel and Schmidt 1991). Although this gecko species is known to be territorial, there are no studies available on the exact territory size. The study population shows realistic density-dependent spatial distribution (Wanger et al. in press) and stop-and-go behavioral patterns (some individuals move around and then stop for some days in a territory before moving on to another; Furrer et al. 2006).

Photo-recognition for individual identification

We used the distinctively different color patterns of each gecko (shining-green ground color with a dorsal pattern of red dots and/or lines) for individual photo recognition, thus avoiding the need for disruptive handling and invasive marking techniques. Photo-recognition has successfully been applied to numerous aquatic and terrestrial species over different periods of time (e.g., Miranda et al. 2005; Grafe et al. 2006; Speed et al. 2007; Van Tienhoven et al. 2007). However, concern was raised about the probability of misidentification resulting from (1) poor photo quality, (2) lack of distinctiveness such as obvious and readily recognizable marks, and (3) a lack of mark stability over time that may lead to an obvious bias when estimating demographic parameters (Stevick et al. 2001). In order to minimize these biases, we firstly only used photographs, where at least two-thirds of the dorsal pattern was clearly visible (i.e., the individual in the picture for identification was filling out the whole picture). Secondly, marks of this gecko species comprise homogeneous red dots on a shiny green background, which are highly distinctive and easily recognizable. And finally, based on the literature (Hallmann et al. 1997), statements from various keepers of this species (T.C. Wanger, pers. comm.), and our own observations (we did not find changes in color pattern in any of the 18 shedding individuals observed and we could successfully re-identify ten geckos that were photographed before released into the exhibit in 2003) we can assume that loss of marks induces little bias, if any. In addition, newly taken pictures were matched independently without any contradictions by two persons to the available pictures. Therefore, although there is the possibility for a potential bias, we consider it to be minimal.

Sampling protocol

For CMR and LT sampling, we used five individual transect lines of 75 m each along the established footpaths that were randomized for CMR in sampling sequence before each sampling session. The combination of all transects after each sampling session resulted in a single transect line of 375 m length and a width of 5 m each side, giving a representative picture of the heterogeneous exhibit area (~39.5% of the total exhibit area was sampled, if the area of water bodies is subtracted; see Fig. 1). After a 1-week pilot study, we carried out 30 and 27 sampling sessions for CRM and LT, respectively, in the following 3 weeks with two people. We chose two time slots from 10 to 12 h and from 15 to 17 h as we found the geckos to be most active during these times in the pilot study. For each person, sampling was interchanged daily between the morning and afternoon session (sampling time 1 h), because we always conducted LT sampling in parallel.

Tested methods

We chose five different methods to estimate population size that all allow heterogeneity in the capture–recapture data: (1) the still commonly used program CAPTURE (Otis et al. 1978) and its implemented selection procedure for choosing the most appropriate model, (2) the recent finite-mixture model of Pledger (2000), (3) an extension based on Pledger's models, using model averaging for estimating population size based on the Akaike's information criterion (AIC; Burnham and Anderson 2001), (4) the accumulation curve method (Kohn et al. 1999), to date used mainly for estimating population sizes from recaptures of genetic markers and for estimation of species richness, and (5) the line transect method implemented in the program DISTANCE (Buckland et al. 2001), which is also based on AIC to choose the most parsimonious model regarding the fit of the detection curve.

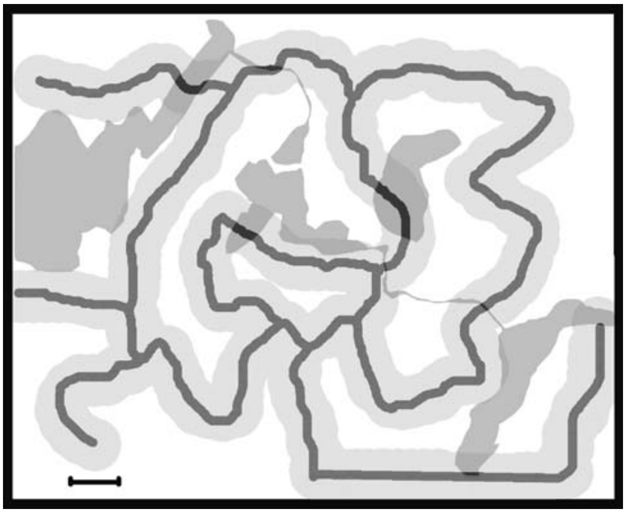

Fig. 1 Map of the Masoala rainforest exhibit showing the transect (*black line*) with transect width (*light grey area* around the transect line) and water bodies (*dark grey areas*). *Scale bar* = 10 m

Each method allows for "fine-tuning" based on additional information (e.g., covariates such as size or information on weather data, to constrain capture rates in CMR or line transect data) that is sometimes available. However, this type of additional information is not always at hand and many non-specialist users are not fully familiar with these possibilities. We, therefore, used only the "raw" methods to estimate population size for further testing of robustness and applicability of these methods to non-professional users and standard data sets.

We used two ways to evaluate the performance of these methods. The first was the CV: the degree to which a set of data points varies relative to the mean. We calculated the CV, which can be used to assess the precision of the sampling technique, for the full data set. It should be below 20% for most management activities (Krebs 1998). A second way to analyze the effectiveness and efficiency of different methods was to use subsets of the data. We used the recapture data from the first 5, 10, 15, 20, and 25 sessions, respectively, and then compared it to the estimate derived from the full data set (30 sessions). This evaluation is based on the assumption that different estimates of principally unbiased methods should result in similar estimates when an extensive dataset is used (see e.g., Brook and Kikkawa 1998).

CMR methods

For the CMR analysis we used the program CAPTURE that includes seven different models to estimate population size in closed populations to deal with heterogeneity in recapture rates. The three possibilities of variation are time-dependence (model M_t), trap response (model M_b) and individual variation (model M_h). We used the implemented selection procedure of CAPTURE, based on a number of simulated datasets, a combination of χ^2 tests, and a regression approach, to choose between models.

Closed CMR model assumptions are as follows: (1) the population is closed—this assumption is not violated as we conducted our study on a captive population. We did not cover the whole area (see map of the enclosure, Fig. 1) and, therefore, were probably not able to capture all individuals. This leads to a population estimate that covers only the individuals which have a principle capture probability above zero (territorial individuals next to the transect and floating individuals). Therefore, the estimate is likely to be negatively biased for the whole population, but as this bias is the same for all tested methods the estimates should be still comparable. Due to a short sampling period (3 weeks) in comparison to the expected live span of the animals (geckos are known to have an exceptional long life span, see Bannock et al. 1999) and the exclusion of newborns, we assume no demographic processes to occur. (2) No loss of marks over the entire sample period—this assumption is not violated as we used photo-recognition to identify individuals (for further details see photo-recognition section above). (3) Marked and unmarked individuals mix completely between sampling sessions—since some individuals are territorial whilst others are not, this assumption is certainly broken; however, dealing with methods that can well account for heterogeneity in the data, violation of this assumption should not cause enormous problems. (4) No marking/trap specific reactions such as trap happiness or trap shyness—we do not need to take this assumption into account, as we did not use physical markings and animals are used to visitors. (5) Catchability does not vary within individuals—presumably this assumption is violated as neither all animals had their territories within the same distance to the transect line nor was the position on the plants equally easily detectable. (6) There is equal catchability of each individual in every sample session—this assumption is not violated. Even though weather conditions change and are strongly correlated with sightings of geckos (Kitchener et al. 1988; Henle 1990; Furrer et al. 2006), they do affect all individuals equally in all sampling sessions. However, assumption (4), (5), and (6) can be relaxed by using the appropriate models M_b, M_h, and M_t, respectively, to account for variation in recapture probability.

Pledger's finite-mixture models

Pledger (2000) suggested additional models that allow for heterogeneity in the capture–recapture probabilities. The basic idea is, that there are unknown groups of animals (mixtures), for example territorial versus non-territorial animals, age, or gender (if sex cannot be determined during capture), that differ in their capture rates. We used the Pledger models (included in the program MARK; White and Burnham 1999) allowing for two mixtures; hence, in these models an additional parameter is estimated that describes the probability of an animal being in mixture one or two. The number of mixtures has to be assumed beforehand, but when testing an additional number of mixtures, we found similar results. When we tested for the effect of age (juvenile vs. adult) and territorial versus floater (i.e., mean distance moved between captures) on capture probability, we found that only age had an almost significant effect ($P = 0.06$) on capture probability. This might explain why two mixtures resulted in the most parsimonious model (i.e., the lowest AIC value).

Model averaging

Estimating population size using only the Kullback–Leibler best model (as inferred from AIC) will inevitablyresult in a biased measure, because model selection uncertainty is not incorporated in the estimate (Burnham and Anderson 2001). To overcome this problem, we used the complete set of candidate models to estimate population size, using model averaging based on AIC weights (essentially, proportional likelihood), indicating the appropriateness of each model (Burnham and Anderson 2001). We used the standard set of Pledger's mixture models, allowing for time-dependence in capture and/or recapture probabilities. This includes the most commonly used models in the capture–recapture framework (models M_0, M_b, M_t, M_h, and their combinations). Some of the models do not allow population size to be estimated unless they are constrained (e.g., M_{tbh}); these were excluded from the model averaging procedure.

Accumulation curve method

The principle of the ACM method is to fit the cumulative number of recorded individuals to the number of sampling sessions (see Colwell and Coddington 1994 for a review). The asymptote of the curve is then an estimation of the total individual number present in the area (Petit and Valiere 2006). Accumulation data can be fitted to various equations. After testing several of these equations and obtaining similar results, we applied the most commonly used one (proposed by Kohn et al. 1999), $y = ax/(b + x)$, where y is the cumulative number of individuals captured until the xth session. a is the quantity of interest, hence the asymptotic number of individuals and b is the half-saturation parameter but has no actual biological interpretation (b is related to the decrease of the slope of the asymptote). A standard non-linear regression-based least-squares was used to estimate a and b. We then used the standard error of the fitted model parameter a to calculate the coefficient of variation (CV) for 30 sampling sessions.

Line transect methods

The main difference between CMR and LT is that there is no need to individually recognize animals in the latter method. As an underlying idea, LT methods estimate the density of animals in a strip by plotting number of sighted animals against distance from the transect line. By fitting a curve to the observed frequencies, the probability of detecting all sampled animals (analogous to the capture rate in CMR methods) per sampled area can be estimated. Specifically, the following assumptions have to be met: (1) Individuals located directly on the transect line are always detected—as the well-established footpaths in the exhibit were used as a sampling line, we

Table 1 Capture–recapture summary of the complete data set

Observation occasion (*j*)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
n (*j*)	9	8	3	6	10	2	9	1	6	4	5	4	6	3	3	4	2	8	7	7	5	6	1	2	1	12	10	4	10	9
M (*j*)	9	13	15	17	22	22	24	24	29	30	31	32	36	37	39	40	40	43	43	47	51	54	55	55	55	58	59	60	63	65
u (*j*)	9	4	2	2	5	0	2	0	5	1	1	1	4	1	2	1	0	3	0	4	4	3	1	0	0	3	1	1	3	2

Individual animals sighted	1×	2×	3×	4×	5×	6×	7×	8×	9×
f (*j*)	31	11	8	5	2	1	4	1	2

For the complete dataset see the Appendix (Supplementary material); *n* (*j*) = animals sighted; *M* (*j*) = cumulative total; *u* (*j*) = newly sighted; *f* (*j*) = frequencies; *j* = observation occasion

can be sure that this assumption is not violated. (2) Animals were detected prior any movement away from their original location due to the observer—this assumption is not violated as geckos in the Masoala exhibit are used to visitors, have a short flight distance (<2 m), and were never approached any closer. (3) Distances are measured correctly—we used a tape-measure to determine radial distances to the nearest centimeter. LT data were analyzed with the program DISTANCE 5.0 (Thomas et al. 2006). To avoid overlooking of geckos located further away, we set the maximum distance to record a gecko to 5 m away from the transect line. To allow for a comparison of the estimates of CMR methods, we chose the reference area to be the complete area of the exhibit (sum of all transects) × 10 m (5 m distance from the line to each side) = 3,750 m^2.

Results

Capture history

Table 1 gives a detailed summary of the complete capture dataset for the CRM (see Appendix (Supplementary material) for the full dataset). Altogether, 167 animals were "captured" (photo-identified), comprising 65 different individuals in total. Average daily number of captures was ~5.6 animals, leading to an average capture probability of ~5% per session. The frequencies of captures for individuals were as follows; 31 animals were captured only once, 11 twice, eight three times, and five were captured four times. The curve then tails off to only single animals captured various times, and two animals captured nine times. Some support for heterogeneity in recapture comes from the observation that new animals were still captured (two to three individuals) in the last few sessions. We can, therefore, assume a high number of animals with a capture probability close to zero. This type of low-recovery system is often associated with difficulties in estimating population size with precision (Pledger 2005). For the LT data, we recorded 47 distances in the complete dataset. The best-fitted model for the detection function was the hazard-rate model with no adjustment term (Table 2).

Table 2 Summary of line transect data

Interval (m)	Observed values	Expected values	Chi-square values
0–1	10	11.33	0.157
1–2	12	11.33	0.039
2–3	12	11.33	0.039
3–4	10	10	0
4–5	3	3	0

Total Chi-square value = 0.2353; *df* = 2. Probability of a greater *Chi-square* value (*P*) = 0.889

Table 3 Population estimates and coefficient of variation for the full data and for pooling of 6 days into one session

Method	Number of sessions	Population estimate	Standard error	Coefficient of variation
Capture	30 (6)	112 (126)	16.8 (15.3)	0.150 (0.121)
Pledger	30 (6)	110 (116)	32.3 (37.5)	0.293 (0.324)
Model averaging	30 (6)	107 (112)	27.9 (36.3)	0.261 (0.32)
ACM	30 (6)	112 (113)	1.0 (2.0)	0.009 (0.018)
LT	27	66	5.95	0.090

Note the similar estimates of population size; pooled data given in brackets

Population estimate ($\hat{N}$) and coefficient of variation

All methods showed similar estimates of population size $\hat{N}$ (107–112 for the different methods, Table 3) for the complete data set (30 sessions) except from the LT method ($\hat{N} = 66$) that differed considerably from CMR methods. We will, therefore, from now on consider 107–112 animals as the reference value for the most likely population size for comparing the four CMR methods. The coefficient of variation is below 0.2 for all methods except for Pledger's finite-mixtures, with and without model averaging. The most precise estimate, with the lowest CV (only around 1% for the full data set), is that of the ACM.

We assessed lumping effects by pooling 6 days into one sampling session because even using 15 sampling sessions as subsets of the total data (see next section) is more than in most field experiments. Pooling slightly reduced the CV for the capture models whilst the CV for the other models was marginally higher. The estimate of $\hat{N}$ did not change much with the capture models showing

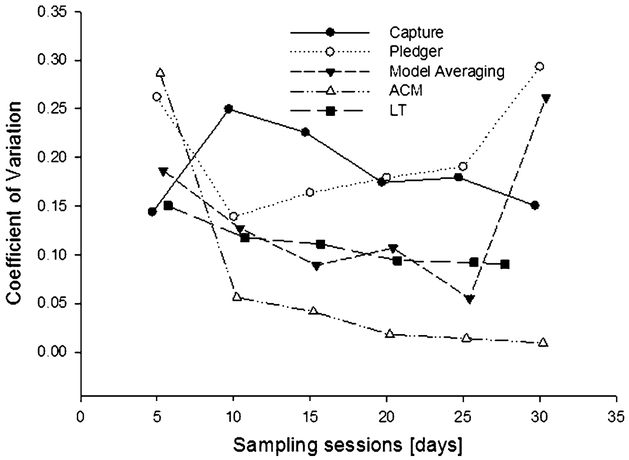

Fig. 2 Coefficient of variation for all tested methods over the time period of sampling. The first, second, third, fourth, and fifth cluster of *dots* represents 5, 10, 15, 20, and 25 subsets of the sampling data, respectively; the sixth cluster represents the full dataset of 30 sampling sessions (27 sampling sessions for LT). The abbreviations *Capture*, *Pledger*, *Model averaging*, *ACM*, and *LT* refer to the models implemented in the program CAPTURE, Pledger's finite-mixture model (Pledger 2000), model averaging based on Akaike's information criterion, the accumulation curve method, and line transect method (*LT*), respectively, that were used to estimate population size

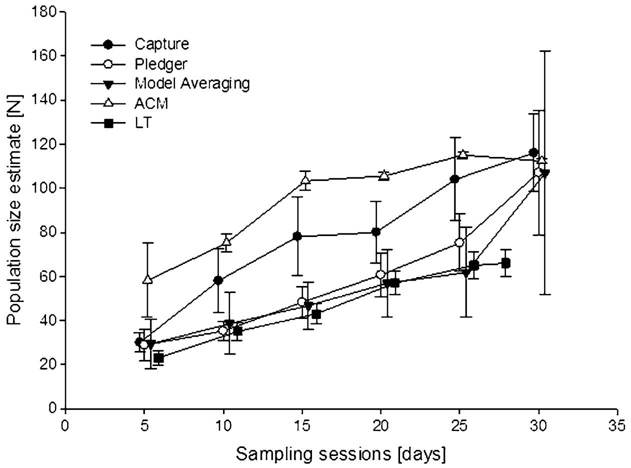

Fig. 3 Comparison of population size estimates of all tested methods. Estimates including *standard error bars* are shown for complete and subsets of the sampling data represented by each cluster of *dots* (for explanation see Fig. 2) over the period of time sampled. Capture, Pledger, Model Averaging, ACM, and LT refer to the methods used to estimate population size in this study (see also Fig. 2)

the largest increase from 112 to 126 for 30 sessions (see Table 3 for a detailed comparison).

Performance of the methods over subsets of data

Coefficient of variation

With an increasing number of sessions, a robust method should deliver a more precise estimate (lower CV). The coefficient of variation decreases with increasing sample size for all methods except for the Pledger finite-mixture models (Fig. 2). The Pledger CV, with or without model averaging, remains steady, or even increases, as more sessions are included in the analysis. The ACM has by far the smallest CV (< 0.01) followed by the LT (< 0.1). The method using CAPTURE falls below the critical 20% level when sampling effort is 20 sessions or more, whereas the LT and ACM are already below this critical value at five and ten sampling sessions, respectively.

Population estimate

It is important to note here that a small CV does not necessarily indicate a valid estimate. It may indicate a faulty estimator; a small CV leads to a small confidence interval which is more likely not to cover the correct population size compared to more conservative methods. For example, the CV of the ACM at session ten is misleadingly narrow, because the resulting 95% confidence interval (66–83), does not include the reference value of 107–112 animals (Fig. 3). In general, however, all of the methods are negatively biased (underestimate population size) with smaller data sets (Fig. 3). With increasing sample size, the ACM is the first to level off at approximately the reference population size, after about 15 sessions. The other methods, however, show a more-or-less constant increase in population estimate with higher sampling effort. CAPTURE and Pledger model do show estimates that are congruent with the ACM after 30 sessions, but, due to their constant increase, give no indication that the correct level of population size has been reached.

The estimates for LT are lowest for all subsamples except from the 25 session subsample. For the full data set, the LT population estimate is considerably lower compared to CRM (66 and 107–112, respectively).

Discussion

Comparing LT and CMR

During the sampling period, line transect data (see Buckland et al. 2001) were also obtained. LT data are different from CMR data in that no marking of individuals takes place. This may be imperative in circumstances where the marking of individuals is practically difficult. An interesting result of our study is the difference in estimates of population size between CMR methods ($\hat{N} = 106–112$) and the LT method ($\hat{N} = 66$), when using the full data set. However, this difference has to be considered with caution and on the basis of the two different approaches these methods use. Although both methods estimate population size, LT estimate density in the first place and, hence, to obtain a population size estimate, density must be multiplied by the reference area (here 3,750 m^2). On the contrary, the CMR data estimates population size of all geckos having a capture probability bigger than zero. As capture effort was restricted to transect lines, animals having their territory far away from the transect line will never be captured. These animals will not be included in the CMR-based population size estimate.

If we use the entire area of the rainforest exhibition as reference area for the LT data, we estimate $\hat{N} = 164$ (confidence interval = 137–197), which is considerably higher than the estimate from the CMR data. If we consider this estimate as valid for the whole exhibit, then this hints that indeed some animals are missed by the CMR methods.

As all CMR methods use the same kind of data, the resulting bias is the same for all CMR methods and, therefore, estimates are comparable. In contrast, bias for the LT is different and, hence, the estimates of both method types are not comparable.

Comparing CMR methods

Reassuringly, when applied to the full 30-session data set, all CMR methods showed similar point estimates of population size. Nevertheless, the methods differed strongly in the width of the confidence interval, sometimes resulting in an unrealistic estimate of precision (e.g., ACM). All methods underestimated population size when the sampling effort was small. Theory suggests that an underestimation of population size is always to be expected if there is heterogeneity in capture probability. When animals are always missing, capture probability of some individuals will drop close to zero (Pledger 2000). In our dataset, this is clearly the case because some new animals are still detected beyond sessions numbered above 25 (recall that this is a closed population, so these animals were all present but unobserved at the first capture session). The simple accumulation curve method gives a good result under these conditions, but does not include the "true" population size before a sampling effort of 25 sessions is reached. The capture models approach the reference value (107–112 animals) more slowly, but also cover the reference value, with their broader confidence intervals, after 25 sessions. Note that the minimum population size is 65 individuals as this is the number of individuals captured and, hence, the Pledger model and model averaging methods do severely underestimate population size with lower sampling effort. Therefore, the best selected model of Pledger and also the Pledger method based on model averaging show only useful estimates when the full data set is used. Model averaging does slightly improve the estimate (lower standard error and smaller confidence interval), but the CV is still higher than 0.2 for the full dataset and, therefore, is not useful in a monitoring context.

Since our sampling period was relatively short, the difference in population estimates with respect to sampling times may also result from homogenization of marked/unmarked individuals. Pooling of the data in fewer blocks of time should then result in different population estimates and CVs—in our case, however, pooling did not change the results.

Depending on the goal of a given population study, we can give different recommendations from this evaluation. The simple ACM should be tested further, ideally with data where the census population size is known, as it is not clear which form of asymptotic curve is linked to which type of capture heterogeneity (Petit and Valiere 2006). If the aim of the wildlife manager had been simply to monitor whether a population remained above its initial state of 60 animals, any of the methods could be used (having accumulating at least 25 sessions). Conversely, if the aim was monitoring population size with the aim of detecting a 10% difference compared to a reference size, then none of the methods except the ACM could be used. Estimates based on the program CAPTURE performed better than the Pledger's models, although the latter are said to perform well in studies with high heterogeneity.

Further rigorous testing of the methods with field data should clearly be undertaken to improve understanding of the context within which each method is most suitable. Our approach of using subsets of data to understand if and when estimates of population size level off should be further explored since it could serve as an alternative route to quantifying confidence in ongoing monitoring estimates. Moreover, we would encourage development of software that combines different methods of estimating population sizes (and also facilitates analysis of subsets of data), to see if/when methods do converge to a common estimate. Such an approach is already underway for the program DISTANCE, where CMR techniques can only be implemented to a limited extent in a mainly line transect oriented "environment" (Buckland et al. 2001 and later Web updates). Simulation of data can assist in this process, but would be less helpful in choosing the right method for a set of field data.

Moreover, it would be interesting to understand the more general reasons for underlying differences in the observed method performances. This, however, requires a lot more model testing (using realistic simulation approaches and alike) as it has been used for simulated data and species richness estimators (e.g., Walther and Moore 2005 and references herein). Note that explaining the overall differences between methods and how heterogeneity is treated between the models does not help much as the overall performance strongly dependent on the nature of the data. We used field data as simulated data often does not reflect real field data (e.g., being based on specific assumption on the distribution of recapture probabilities or enormous sampling effort). Assessment of general performances is not a qualitative task on itself, as most of the assumptions and their violations do bias the estimate in opposite directions and, therefore, cannot be solved by thoughtful arguments alone. Future studies using field data will facilitate a meta-analysis to provide more general insights to evaluate method performance and, hence, allow for more user-friendly recommendations. However, at present, such an analysis is still lacking due to unavailability of comprehensive sets of field data (i.e., "over-sampled" and with a roughly known population size).

We argue that scientists focused on the development of methods of statistical inference should endeavor to simplify, where possible, the application of existing techniques to give wildlife managers user-friendly step-by-step recommendations on how population estimates should be performed. We are fully aware that further simplification carries the danger of using inappropriate methods for the wrong type of data, but the negligence of scientists to simplify the application of their methods has led to a wide gap between advancing theory and methods, and its actual use by most wildlife managers.

Acknowledgments We thank Zurich Zoo for providing logistic support during the data collection. M. Bauert allowed us to work in the Masoala exhibit and T. Zellweger and his animal keeper team provided us with all details about the exhibit. All work on the gecko population was conducted with permission of Zurich Zoo and in accordance with their approved guidelines. We also thank Lochran Traill and two anonymous reviewers for their insightful comments on the manuscript.

References

Bannock CA, Whitaker AH, Hickling GJ (1999) Extreme longevity of the common gecko (*Hoplodactylus maculatus*) on Motunau Island, Canterbury, New Zealand. N Z J Ecol 23:101–103

Brook BW, Kikkawa J (1998) Examining threats faced by island birds: a population viability analysis on the Capricorn silvereye using long-term data. J Appl Ecol 35:491–503

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling. Oxford University Press, Oxford

Burnham KP, Anderson DR (2001) Kullback–Leibler information as a basis for strong inference in ecological studies. Wildl Res 28:111–119

Chao A (2001) An overview of closed capture–recapture models. J Agric Biol Environ Stat 6:158–175

Chao A, Chu WT, Hsu CH (2000) Capture-recapture when time and behavioural response affect capture probabilities. Biometrics 56:427–433

Coull BA, Agresti A (1999) The use of mixed logit models to reflect heterogeneity in capture–recapture studies. Biometrics 55:294–301

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Phils Trans R Soc Lond B Biol 345:101–118

Eggert LS, Eggert JA, Woodruff DS (2003) Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Mol Ecol 12:1389–1402

Freilich JE, Burnham KP, Collins CM, Garry CA (2000) Factors affecting population assessment of desert tortoises. Conserv Biol 14:1479–1489

Furrer SC, Jaag K, von Stockar S, Rübel A (2006) First experiences with free-ranging giant day geckos (*Phelsuma madagascariensis grandis*, Gray 1870) in the Masoala rainforest exhibit in Zurich Zoo, Switzerland. Zoo Biol 25:409–415

Grafe TU, Wanger TC, Dehling JM, Proksch S, Braasch T (2006) Use of individual markings in the study of the foot-flagging frog, *Staurois guttatus*. Scientia Bruneiana 7:3–8

Gruber B, Henle K (2004) Linking habitat structure and orientation in an arboreal species *Gehyra variegata* (Gekkonidae). Oikos 107:406–414

Hallmann G, Krüger J, Trautmann G (1997) Faszinierende Taggeckos. Die Gattung Phelsuma. Natur und Tier Verlag, p 232

Henkel FW, Schmidt W (1991) Geckos: Biologie. Haltung und Zucht. Ulmer Verlag, Stuttgart, p 224

Henle K (1990) Population ecology and life history of the arboreal gecko *Gehyra variegata* in arid Australia. Herpetol Monogr 4:30–60

Kohn MH, York EC, Kamradt DA, Haugt G, Sauvajot RM, Wayne RK (1999) Estimating population size by genotyping faeces. Philos Trans R Soc Lond B Biol 266:657–663

Kitchener DJ, How RA, Dell J (1988) Biology of *Oedura reticulata* and *Gehyra variegata* (Gekkonidae) in an isolated woodland of Western Australia. J Herpetol 22:401–412

Krebs CJ (1998) Ecological methodology, 2nd edn. Benjamin Cummings, Menlo Park, California, p 624

Lukacs PM, Burnham KP (2005) Review of capture–recapture methods applicable to noninvasive genetic sampling. Mol Ecol 14:3909–3919

Miranda T, Ebner M, Solé M, Kwet A (2005) Estemativa populacional de *Psedis cardosoi* (Anura, Hylidae), com emprego de métoco fotográfico para reconhecimento individual. Biociências 13:49–54

Otis DL, Burnham KP, White GC, Anderson DR (1978) Statistical inference from capture data on closed animal populations. Wildl Monogr 62:1–135

Palsbøll PJ, Allen J, Bérubé M (1997) Genetic tagging of humpback whales. Nature 388:767–769

Petit E, Valiere N (2006) Estimating population size with noninvasive capture–mark–recapture data. Conserv Biol 20:1062–1073

Pledger S (2000) Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56:434–444

Pledger S (2005) The performance of mixture models in heterogeneous closed population capture–recapture. Biometrics 61:868–873

Savarie PJ, Shivik JA, White GC, Hurley JC, Clark L (2001) Use of Acetaminophen for large scale control of brown treesnakes. J Wildl Manage 65:356–365

Schaub M, Pradel R, Jenni L, Lebreton J-D (2001) Migrating birds stop over longer than usually thought: an improved capture–recapture analysis. Ecology 82:852–859

Schwarz CJ, Seber GAF (1999) Estimating animal abundance: review III. Stat Sci 14:427–456

Silver SC, Ostro LET, Marsh LK, Maffei L, Noss AJ, Kelly MJ, Wallace RB, Gómez H, Ayala G (2004) The use of camera traps for estimating jaguar *Panthera onca* abundance and density using capture/recapture analysis. Oryx 38:148–154

Speed CW, Meekan MG, Bradshaw CJA (2007) Spot the match—wildlife photo-identification using information theory. Front Zool 4:2

Stevick PT, Palsbøll PJ, Smith TD, Bravington MV, Hammond PS (2001) Errors in identification using natural markings: rates, sources, and effects on capture-recapture estimates of abundance. Can J Fish Aquat Sci 58:1861–1870

Thomas L, Laake JL, Strindberg S, Marques FFC, Buckland ST, Borchers DL, Anderson DR, Burnham KP, Hedley SL, Pollard JH, Bishop JRB, Marques TA (2006) Distance 5.0. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/

Tyrell CL, Cree A, Towns DR (2000) Variation in reproduction and condition of northern tuatara (*Sphenodon punctatus punctatus*) in the presence and absence of kiore. Sci Conserv 153:5–42

Van Tienhoven AM, den Hartog JC, Reijns RA, Peddemors VM (2007) A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark *Carcharias taurus*. J Appl Ecol 44:273–280

Walther BA, Moore JL (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography 28:815–829

Wanger TC, Motzke I, Furrer SC, Brook BW, Gruber B (in press) Movement patterns and habitat selection of the gaint day gecko (*Phelsuma madagascariensis grandis*) in the Masoala rainforest exhibit, Zurich Zoo. Salamandra, p 13, Fig. 4

White GC, Burnham KP (1999) Program MARK: survival estimation form populations of marked animals. Bird Study 46:120–138

Williams BK, Nichols JD, Conroy MJ (2002) Analysis and management of animal populations. Academic Press , San Diego

Wood KV, Nichols JD, Percival HF, Hines JE (1998) Size–sex variation in survival rates and abundance of pig frogs, *Rana grylio*, in northern Florida wetlands. J Herpetol 32:527–535

Yip PSF, Xi LQ, Chao A, Hwang W-H (2000) Estimating the population size with a behavioral response in capture–recapture experiment. Environ Ecol Stat 7:405–414