J Ornithol (2012) 152 (Suppl 2):S371–S380
DOI 10.1007/s10336-010-0542-8

EURING PROCEEDINGS

Heterogeneity in detection probability along the breeding season in Black-legged Kittiwakes: implications for sampling design

**Thierry Chambert · Deborah Pardo · Rémi Choquet ·
Vincent Staszewski · Karen D. McCoy ·
Torkild Tveraa · Thierry Boulinier**

Received: 13 September 2009 / Revised: 5 February 2010 / Accepted: 26 May 2010 / Published online: 26 June 2010

Abstract In wild animal population studies, capture heterogeneity is likely to be prevalent and can reduce the accuracy of vital rate estimates. Here, we test how individual detection probabilities vary through the breeding season in a population of a cliff-nesting colonial seabird, the Black-legged Kittiwake (*Rissa tridactyla*). Specifically, we expected detection probability to be affected by changes in brooding behavior and nest attendance associated with the breeding phenology and the local breeding performance of individuals. As predicted, we found that strong heterogeneities in detection probability can occur in relation to the breeding performance of individuals, the breeding performance of their neighbors, and the timing of surveys. Detection probability is highest and most homogeneous at the beginning of the breeding season. Later in the season, it is lower and can vary dramatically among groups of breeding individuals. A simulation approach was used to assess the implications of these results for the performance (bias and precision) of different study designs. Clearly, investing sampling effort early in the season is an efficient way to improve the accuracy of parameter estimates in this species. Our findings stress the importance of establishing study designs that take into account the population and behavioral ecology of the focal species.

Keywords Study design · Nest attendance · Capture–mark–recapture · Prospecting behavior · Survival rate · *Rissa tridactyla*

Introduction

Reliable estimates of key population and community dynamic parameters are important not only for conservation and wildlife management purposes but also for basic studies in evolutionary ecology (Yoccoz et al. 2001). In many cases, capture–mark–recapture (CMR) approaches are used to account for the fact that not all individuals are detected at each sampling occasion (Lebreton et al. 1992; Williams et al. 2002). Two processes play a key role in obtaining accurate estimates, and thus drawing effective inferences, in CMR studies, (1) the sampling design applied in the field, and (2) the underlying statistical estimation model (Kendall et al. 2009). Detailed attention has been given to statistical modeling (Pollock et al. 1990; Lebreton et al. 1992), but design issues are sometimes neglected or taken for granted (but see Williams et al. 2002). For instance, in seasonally breeding species, data from series of capture–recapture/resighting occasions are often lumped together and little information is then

---

Communicated by W.L. Kendall.

T. Chambert (✉) · D. Pardo · R. Choquet · T. Boulinier
CEFE, CNRS UMR 5175, 1919 Route de Mende,
34293 Montpellier, France
e-mail: thierrychambert@aliceadsl.fr;
Thierry.Chambert@cefe.cnrs.fr

Present Address:
D. Pardo
CEBC-CNRS Villiers-en-Bois, 79360 Beauvoir-sur- Niort,
France

V. Staszewski
School of Biological Sciences, University of Edinburgh,
Edinburgh, UK

K. D. McCoy
GEMI UMR 2724, CNRS-IRD, Montpellier, France

T. Tveraa
Department of Arctic Ecology, NINA, Tromsø, Norway

provided on the within-season timing of sampling, despite its potential importance.

The quantity and quality of information gathered in the field are strongly affected by the imperfect detection of individually marked animals (Clobert et al. 1993). Appropriate study designs can enhance the accuracy of estimates in two ways: (1) by increasing detection probability or (2) by reducing heterogeneity across individuals (Sanz-Aguilar et al. 2010). The first way increases the quantity of available information and thus the precision of estimates. The second way reduces potential bias due to the violation of the homogeneity assumption (Buckland 1982; Pradel et al. 1997; Prévot-Julliard et al. 1998; Crespin et al. 2008). Therefore capture–recapture study designs should ideally allocate most capture effort to occasions when the rate of detection is high, and the homogeneity assumption is respected.

Heterogeneity in detection probability is recognized as a severe problem for estimating population size (Gilbert 1973; Hwang and Chao 1995), but because early work suggested that its effect on survival estimation was less important (Carothers 1973, 1979), it has been relatively neglected. Heterogeneity in detection probability is known to occur in many field studies and may be due to both (1) extrinsic factors, induced by the study itself, and (2) intrinsic factors, which reflect features of the population (Crespin et al. 2008). Extrinsic factors are commonly invoked to explain heterogeneity in detection probability, but intrinsic factors, such as age (Lebreton et al. 1992; Davis et al. 2003; Crespin et al. 2006), sex (Tavecchia et al. 2002), breeding status (Clobert 1995; Viallefont et al. 1995; Grosbois and Thompson 2005), social status (Summerlin and Wolfe 1973; Drickamer et al. 1999; Ogutu et al. 2006), infectious status (Senar and Conroy 2004; Faustino et al. 2004), and environment (Pollock et al. 2006), have also been described. In our case, we expected two intrinsic factors to be responsible for heterogeneity in detection probability: brooding and prospecting behaviors. Those two behavioral traits are linked to the breeding status of individual birds.

Kittiwakes, and more generally seabirds, are widely studied in long-term programs and provide useful data to test relationships between population dynamic parameters and environmental changes (Frederiksen et al. 2005). Resighting of banded individuals is performed in colonies during the breeding season. When engaged in breeding activities, individuals are detectable when (1) they are attending their nest site and (2) their legs are visible. The stage of the breeding season is thus likely to affect their detectability (e.g., birds brooding eggs have their legs hidden most of the time), but possibly in interaction with the breeding performance of the marked birds and that of their neighbors. For instance, early interruption of brooding or massive local failures may affect behavior on the nest site and attendance (Boulinier et al. 2008).

In this paper, we investigate how detection probability varies along the breeding season in a population of Black-legged Kittiwakes (*Rissa tridactyla*). We predicted that, at the beginning of the season, when birds are settling and building nests, detection probability would be homogeneous among individuals and high because all birds partake in the same activity and their legs are not hidden in the nest. Later in the season, most birds have laid eggs and, as long as they continue to brood their eggs, they are likely to be difficult to detect. Individuals which lose their eggs, notably because of predation, will stop brooding and may become more detectable again. Individuals that successfully hatch a chick may remain difficult to detect until the chick gets older or disappears from the nest. Finally, the local success of cliffs may vary dramatically and this may affect nest site attendance and thus detection probability late in the season. CMR model selection was used to examine these predictions. The results of these analyses were then used in a simulation approach to assess their implications for the performance (bias and precision) of different study designs.

Methods

Data collection

Resighting of marked birds is conducted each year on Hornøya, an island in Northern Norway (70°22′ N, 31°10′ E), where approximately 21,000 pairs of Kittiwakes breed (Anker-Nilssen et al. 2000). Kittiwakes are colonial birds which reproduce on cliffs and show high breeding site fidelity. Since 1998, a sample of cliffs hosting about 1,000 Kittiwake nests is surveyed each year during the breeding season. About 50 new birds are marked each year with a unique combination of 5 colored bands fixed on both legs that allow their individual identification using binoculars. Resighting of marked birds is conducted every 3 days at each breeding cliff.

In this paper, we used the 2006 intra-annual recapture data. In this year, 330 individual Kittiwakes were resighted in our surveyed plots. When a marked individual is seen, its location (i.e., cliff and nest site) and activity (i.e., displayed behavior) are recorded. Kittiwakes that are engaged in breeding activities display high attendance at their nest within the breeding season. The regular nest site surveys in 2006 allowed us to assign specific breeding sites to 269 of the 330 marked birds recorded that year (i.e., 81.5%). The remaining 61 birds, which were not assigned a nest because they most likely did not breed among any of the study plots, were removed from the data used to run CMR analyses.

Analyses

Field dataset

To estimate the detection rate of marked birds along the breeding season, we used capture–recapture models. Annual survival is high in adult Kittiwakes, ranging from 0.801 to 0.933 (Frederiksen et al. 2005) and is about 0.88–0.92 on Hornøya (Barrett 2001; Sandvik et al. 2005); thus, we expected mortality to be very low within the breeding season, which lasts 3 months. However, as we wanted to avoid potential bias in detection rate estimates due to mortality (or apparent mortality), we used open population models.

To test for differences in detectability due to the reproductive status of individuals, we identified three groups of individuals based on their breeding success and the local success level of the cliff on which they bred (Fig. 1). Because we expected failure at the egg stage to be an important factor affecting detection probability, we considered an individual to be successful if it started to rear a chick that season. We considered a cliff successful if at least 20% of its surveyed nests were successful (i.e., nests detected with a chick during the season). Failed cliffs were those with a nest success of less than 20%. The three groups constituted are thus: group SS (successful individuals that bred on successful cliffs); group FS (failed individuals on successful cliffs); and group FF (failed individuals on failed cliffs, but we actually also included in this group the few banded individuals that hatched a chick on failed cliffs). We included both successful and failed individuals from failed cliffs in the FF group because there are very few successful individuals on unsuccessful cliffs (by definition) and because in 2006 all ‘successful’ individuals on failed cliffs lost their chick well before the end of the season. Once their chick was lost, those birds exhibited the same behavior as unsuccessful birds from failed cliffs (i.e., very low nest attendance).

We also divided the sampling season into three periods corresponding to the three major stages of the breeding season: (1) the nest construction period (NCP) at the beginning of the season when birds are building nests; (2) the brooding period (BP) in the middle of the season when some birds are brooding eggs; and (3) the rearing period (RP) at the end of the season when adults are provisioning young. The three periods were approximately equal in duration that year.

To test our predictions about variation in detectability over the season, we first used a model selection procedure (Lebreton et al. 1992), with two successive steps. In the first step, we built a set of competing models in which survival (Φ) and detection probabilities (p) could: (1) vary or be constant over time, and (2) vary or be constant among groups. For time variation of p, we considered two scales: either variation at every occasion (variation within and among periods), referred as “time effect”, or variation among periods only, referred as “period effect”. The

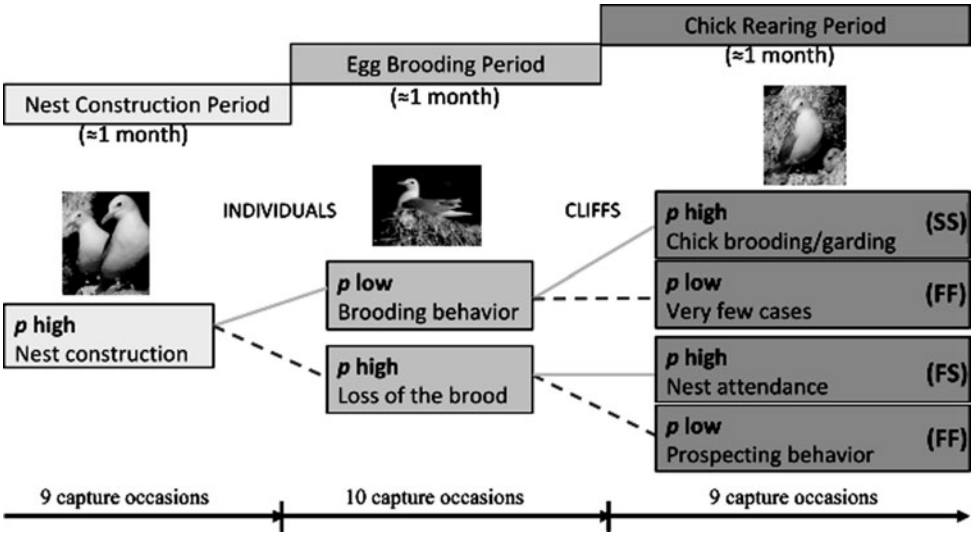

Fig. 1 Illustration of the three main periods of the breeding season of the Black-legged Kittiwake *(Rissa tridactyla)*, with the corresponding breeding behaviors and expected detection probabilities p. Detection probabilities p were expected to vary in time with the reproductive performance of individuals and that of their cliff neighbors (success: *plain line;* or failure: *dashed line*). The *arrow* at the *bottom* of the figure represents the sampling design (number of resighting occasions among the three time periods) used to sample the marked birds during the 2006 reproductive season

"period effect" is a restrictive case of the time effect, because it assumes constancy among occasions within a period. In the second step, we started from the most relevant model of the first step and built a set of new competing models to test if within each period we could identify a group effect or not, and if some groups could be collapsed (indicating that they have the same detection rate for this period, i.e., homogeneity). The most complex model of both sets included time and group effects on both survival and detection probabilities [model $\Phi(g \times t)$, $p(g \times t)$]. We tested whether this model fit the data using program RELEASE (as implemented in MARK; Burnham et al. 1987; White and Burnham 1999). This program evaluates the goodness-of-fit (GOF) of the full CJS model for each group. Analyses were conducted with the software MARK (White and Burnham 1999). The Akaike's information criterion corrected for small sample size (AICc) (Hurvich and Tsai 1989) was used to evaluate the relative support in the data for models in our a priori set (Burnham and Anderson 2002). We then estimated detection probabilities from the most relevant models.

Simulated datasets

We also developed an approach using simulated data. The simulations quantified differences in the accuracy of survival estimates when the sampling effort and the distribution of this effort varied among the three periods of the breeding season (see the legend of Fig. 4 for the different sampling designs simulated). For this purpose, the inter-annual survival rate of a virtual population of 300 Kittiwakes was set to 0.9 and we fixed a relevant detection probability, based on the results from the field data, for each group and period combination. Once this intra-annual pattern was defined, we repeated it for 10 years, the values for Φ and p remaining constant among years. In order to obtain the inter-year capture–recapture histories which are often used in actual multi-year studies, the capture histories of each individual were summarized in series of 1 and 0 corresponding to whether each individual was captured or not at least once over the secondary sampling occasions within a year (i.e., an individual was either "seen" or "not seen" a given year). Simulated data where analyzed with the E-Surge software (Choquet et al. 2009b) using a model that assumed constancy of both parameters [$\Phi(.)$, $p(.)$] for all tested sampling designs and efforts. The relative bias [$(E(\Phi) - \Phi)/\Phi$] and the root mean squared error (RMSE) of the survival estimates were calculated to reflect respectively the bias and the precision in the estimation of that parameter. The fit of the CJS model with full time variation was verified using the U-Care software (Choquet et al. 2009a). The coefficient of overdispersion ($\hat{c}$) was calculated in order to compare how well the model fit could be assessed with simulated data from study designs tested with different sampling effort.

Results

Field dataset

Test of time and group effects

The most complex model in our model set [model $\Phi(g \times t)$, $p(g \times t)$] fit the data ($\chi^2 = 194.44$, $df = 198$, $p = 0.56$). In the first set of competing models, the top-ranked model ($\Phi(g)$, $p(g \times t)$, Table 1) included a group effect for Φ and group and time effects for p (AICc weight = 0.536). However, the second ranked model, which assumes survival to be the same among groups $\Phi(.)$,$p(g \times t)$, was almost as well supported (ΔAIC = 0.29, AICc weight = 0.463). No other model was supported by the data (ΔAIC > 14.21). Estimates of p were nearly identical for the two best models, indicating that they were robust to the way survival was modeled and averaged estimates were thus not needed.

Estimated survival probability was high for both models: 0.995 in the model with equal survival among groups and 0.996 for group SS, 0.990 for group FS, and 0.972 for group FF in the model with group-specificity. As expected, variation in the number of individuals detected along the season is thus mainly due to variation in detection probabilities (Fig. 2).

We found no support for a model in which p varies only between the three periods (ΔAIC = 108). This result does not exclude an effect of periods on p, but rather indicates that periods cannot be considered as the only source of variation of p during the breeding season. In other words, the value of p is not constant within periods (Fig. 2).

Test of homogeneity of detection probability among groups within each period

We used a model with constant Φ to avoid having group specific models over-estimating the value of p for group FF (due to an underestimation of apparent Φ late in the season, $\Phi = 0.89$ instead of 0.97). Most individuals of group FF were not resighted again during the last period of the season. This is unlikely to be due to mortality, but rather because failed breeders from failed cliffs did not attend their nest site (i.e., permanent emigration at the within-season scale).

There were four possible group effects for each period: (1) p is different for all groups (g); (2) p is equal between two groups (SS = FS, FS = FF, SS = FF); (3) p is equal among the three groups (SS = FS = FF); and (4) p is

Table 1 Results of model selection testing for variation in survival (Φ) and detection probability (*p*). Each parameter can either be constant (.) or vary among groups (*g*), among periods (PER), or over all occasions (*t*, i.e., variation within and among periods). We tried all possible combinations

Model	Delta AICc	AICc weight	Model relative likelihood	Num. Par
Φ(*g*), *p*(*g* × *t*)	0.0	0.536	1.000	84
Φ(.), *p*(*g* × *t*)	0.3	0.463	0.864	82
Φ(*g*), *p*(*t*)	14.2	0.000	0.001	30
Φ(*t*), *p*(*g* × *t*)	32.5	0.000	0.000	107
Φ(*g* × *t*), *p*(*g* × *t*)	99.8	0.000	0.000	142
Φ(*g*), *p*(*g* × *PER*)	108.2	0.000	0.000	11
Φ(.), *p*(*g* × *PER*)	109.6	0.000	0.000	9
Φ(*g*), *p*(*g*)	898.9	0.000	0.000	6
Φ(*g*), *p*(.)	911.0	0.000	0.000	4

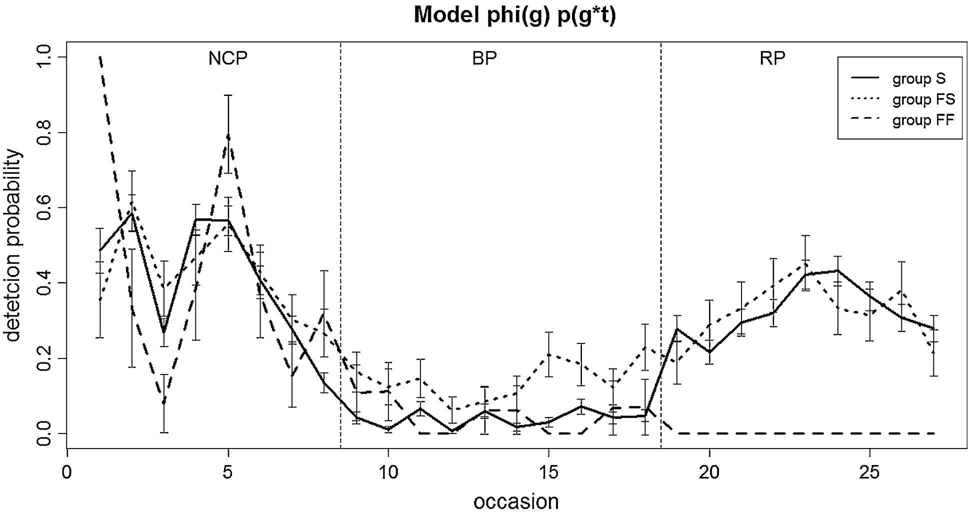

Fig. 2 Variation of *p* along the breeding season for the three groups based on estimates from model Φ(*g*), *p*(*g* × *t*). *Vertical bars* at each occasion represent the standard error of the estimates

different for the three groups, but the difference is constant over occasions (i.e., the group effect is additive (*g* + *t*)). We initially tested all four possibilities for the first period while leaving the group effect (*g*) for the second and third periods. We then kept the structure of the model with the lowest AIC to test several relevant possibilities for the second period, and finally for the last period (Table 2).

The best model structure (AICc weight = 0.995) contained an equal *p* for groups SS and FS in the NCP, different *p* for each group (*g*) in the BP, and equal *p* for groups SS and FS in the RP (model I, Tables 2 and 3). Variations in *p* along the breeding season based on estimates from this model are shown in Fig. 3. As expected, the beginning of the season shows high values of *p* for all three groups. Detection probability ranged from 0.17 to 0.59 for groups SS and FS (mean = 0.40, CI = 0.38, 0.43) and from 0.08 to 1.0 for group FF (mean = 0.35, CI = 0.27, 0.44). However, this period is also the most variable in terms of *p* values, especially for group FF. In the middle of the season, *p* drops for all three groups, but is highest for group FS. As expected, unsuccessful individuals on successful cliffs are more detectable than successful ones during this period (mean = 0.04, CI = 0.03, 0.05 for group SS; and mean = 0.14, CI = 0.11, 0.18 for group FS). In the last period, *p* is very different between successful and failed cliffs. Indeed *p* is relatively high (mean = 0.32, CI = 0.30, 0.35) for groups SS and FS (successful cliffs), while it is zero for group FF (unsuccessful cliffs). These results also highlight that the heterogeneity of *p* among groups is lowest at the start of the breeding season, that is, during the NCP (see estimates in Fig. 3).

Simulated dataset

Simulation results underline the importance of those obtained with the field data; the best estimates are obtained when data collection includes the NCP (Fig. 4). As long as there is at least one sampling occasion during the NCP, precise estimates of survival can be obtained with

Table 2 Set of competing models used to test for the homogeneity among groups within each period

Model	NCP	BP	RP
A	g	g	g
B	SS = FS = FF	g	g
C	g + t	g	g
D	SS = FS	g	g
E	FS = FF	g	g
F	SS = FF	g	g
G	SS = FS	FS = FF	g
H	SS = FS	SS = FS = FF	g
I	SS = FS	g	SS = FS
J	SS = FS	g	SS = FS = FF

In periods containing g, *p* differs among the three groups; for SS = FS, *p* is equal between groups SS and FS; for FS = FF, *p* is equal between groups FS and FF; for SS = FF, *p* is equal between groups SS and FF; and for SS = FS = FF, *p* is equal among the three groups (i.e., homogeneity of detection within the population). Within each period, *p* varies among occasions independently from the group effect, except for $g + t$ (group effect is constant over time). Model A is equivalent to model Φ(.), $p(g \times t)$ from Table 1

NCP Nest construction period, *BP* brooding period, *RP* rearing period

Table 3 Results of model selection for the test of homogeneity in detection probabilities among groups within each period of the breeding season

Model	Delta AICc	AICc weight	Model relative likelihood	Num. par.
I	0.0	0.995	1.000	65
D	11.4	0.003	0.003	74
B	13.0	0.002	0.002	66
C	16.2	0.000	0.000	68
G	16.8	0.000	0.000	64
A	19.7	0.000	0.000	82
E	20.7	0.000	0.000	74
F	20.9	0.000	0.000	74
H	47.0	0.000	0.000	54
J	63.0	0.000	0.000	56

See Table 2 for model details

relatively low bias. Only four study designs exceeded a bias threshold of 0.01 (i.e., high values of relative bias according to Carothers 1979). These four designs are those when sampling only occurred during the brooding and rearing periods, periods when high heterogeneity in *p* has been identified (Figs. 2 and 3).

When we conducted GOF analyses on the general model Φ(t), *p*(*t*) with simulated data from the different study designs tested, we found a $\hat{c} > 5$ when sampling involved many occasions during periods when detection probability *p* is low and heterogeneous among individuals, i.e., late in the season (Fig. 5). This is indicative of a problem within the model structure, in this case due to heterogeneity in *p*. However, for a comparable timing, but with only three occasions, the general model seems to fit the data, despite the fact that our previous results showed a bias in survival estimates in these cases. Therefore, with only few sampling occasions late in the season, the amount of information available is not sufficient to detect heterogeneity in *p* with GOF tests. This could be problematic as, in this situation, one might confidently (based on GOF value) calculate survival estimates which are in fact biased.

Discussion

Patterns of variation in detection probability along the breeding season

Our results highlight that, as expected, dramatic changes in detection probability can occur within a breeding season when surveying a marked population of cliff-nesting species like the Black-legged Kittiwake. More specifically, individuals showed changes in detection probability as a function of the timing of sampling, but also in interaction with their own breeding performance and that of their cliff neighbors. When birds are settling and nests are being built, color-banded individuals have a high probability of being detected because they spend time attending their future nest site and because their legs are not yet hidden in the nest cup. During this period, detection probability is relatively homogeneous among individuals, but can vary dramatically between sampling occasions. During the brooding period, unsuccessful birds are more detectable than successful ones. This difference is certainly due to the brooding behavior of the latter, making it more difficult to see their bands and thus to detect their presence. This difference is not as strong as could be expected if only individuals in the successful group were brooding during this period. As individuals were grouped a posteriori according to their level of breeding performance (and as chick rearing was used as a criterion of success), many of the so-called ‘unsuccessful birds’ may have been brooding eggs for part of the brooding period, making their probability of being detected lower during this period compared to failed individuals standing on their nest. At the end of the season, the difference in detectability among groups depends on the success of the entire cliff on which an individual breeds: the detection rate is relatively high on successful cliffs, whereas it is zero on cliffs with very high rates of failure. This strong difference is certainly due to the fact that birds from failed cliffs spend a great proportion of their time outside their cliff, either foraging or displaying social behaviors like prospecting or aggregating

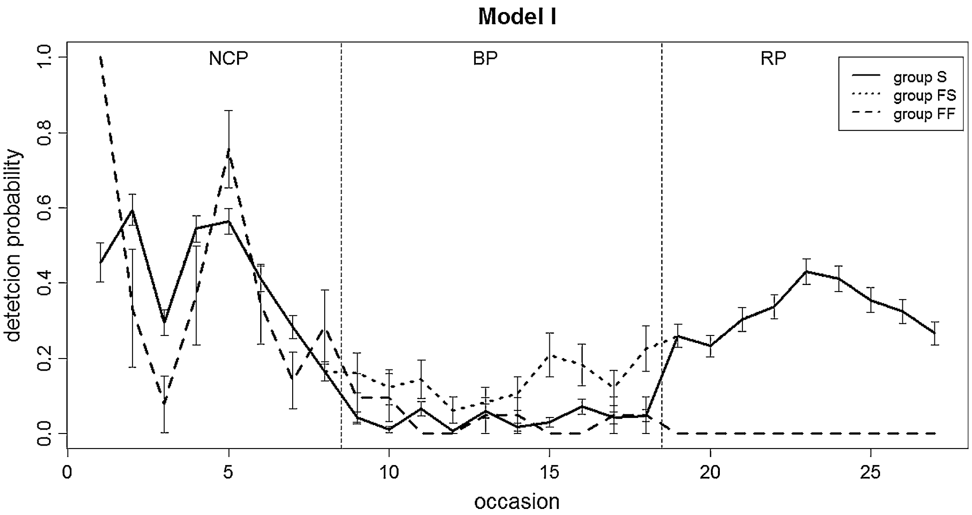

Fig. 3 Variation of p along the breeding season for the three groups of individuals based on estimates from the best-ranked model (model I: see Table 2 for model details). *Vertical bars* represent standard error of the estimates

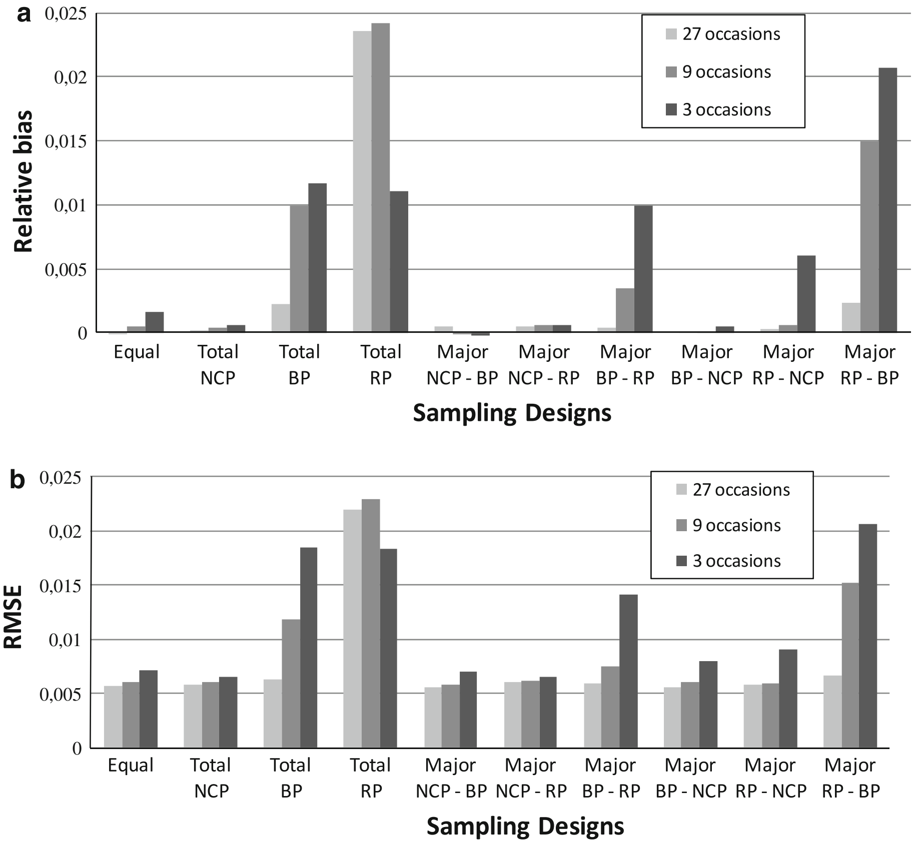

Fig. 4 Results from the simulations showing differences in terms of the **a** relative bias and **b** precision (RMSE) for the different study designs tested. For each design, three sampling efforts are considered: 27, 9 or 3 capture occasions per year. The 10 study designs vary in the division of effort across the different periods of the breeding season: *Equal* means that 1/3 of the total effort is devoted to each period; *Total NCP* corresponds to putting all the effort in the nest construction period, whereas *Total BP* and *Total RP* are, respectively, putting all the effort into the brooding and rearing periods; *Major* means that 2/3 of the sampling effort is allocated to the first period (NCP-BP-RP) cited and the 1/3 left to the second period cited. *NCP* Nest construction period, *BP* brooding period, *RP* rearing period

in 'clubs'. Prospecting behaviors are predicted to be especially beneficial to failed breeders as the end of the breeding season corresponds to the time when reliable information is available about the success of other potential breeding cliffs that the birds could use in future reproduction (Boulinier et al. 1996, Danchin et al. 1998). Low attendance on failed nests on failed cliffs during the chick rearing period has indeed been reported in an experimental study of breeding habitat selection in Kittiwakes, in which it was related to the proportion of individuals that returned to the breeding cliff the year after (Boulinier et al. 2002, 2008). Conversely, high attendance on failed nests on successful cliffs was related to a high fidelity to the breeding patch. This highlights the need to consider

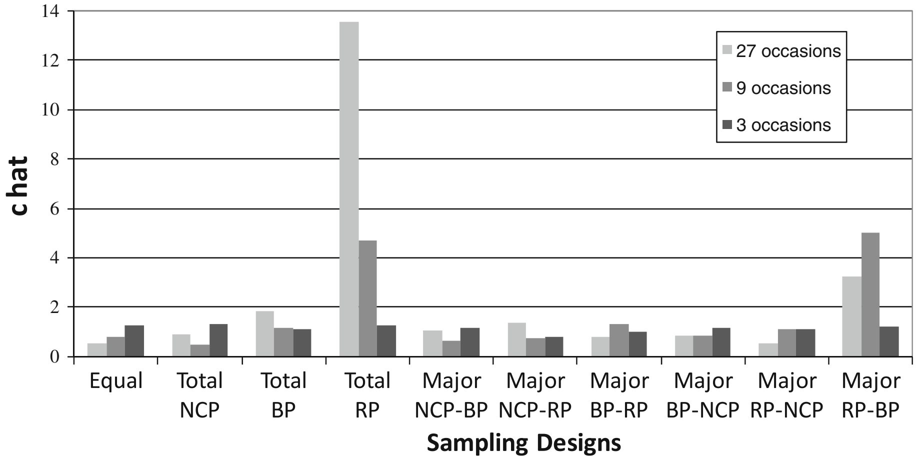

Fig. 5 Results from goodness of fit analyses (GOF) with the CJS general model. We show the values of the overdispersion factors (ĉ) for all sampling designs tested

breeding habitat selection behaviors in such a species when designing field studies for estimating demographic parameters and understanding population dynamics. In particular, our results emphasize that in Kittiwakes there may be strong heterogeneity in detection probabilities among individuals during the two last periods of the season (brooding and chick rearing periods) and that this is certainly linked to changes in the behavior of individuals as the breeding season advances. An extension of our work would be comparing inter-annual variation in detection probabilities among regularly successful cliffs and repeatedly unsuccessful cliffs, with the expectation that in the latter detection should decrease over time.

Variables linked with reproduction, as well as the breeding status itself, have been previously considered as a factor affecting “catchability” (Clobert 1995; Nichols and Kendall 1995; Viallefont et al. 1995; Cézilly et al. 1996). However, our study is among the first to explore how interactions among these types of intrinsic and extrinsic factors (see also Sanz-Aguilar et al. 2010) affect heterogeneity in detectability within a sampling season. This information is useful from both methodological and biological points of view. For instance, it can allow one to directly model heterogeneity in detectability as a function of the breeding status of individuals, rather than to create hidden classes of individuals (see Mixture models in Pledger et al. 2003), especially as we found that the heterogeneity in detection probability was temporally dynamic and not constant over time, as is often assumed (Pledger et al. 2003; Crespin et al. 2008).

Implications of within-season variation in detection probability for study design

The exploration of factors affecting detectability within-seasons can be particularly informative for optimizing population monitoring programs. An efficient study design should attempt to invest most effort in sampling occasions when detection rates are expected to be high and when the homogeneity assumption is verified. Sometimes tradeoffs between reducing heterogeneity and increasing *p* can exist. For example, if *p* and heterogeneity are low, it can be optimal to seek to increase *p* even if it also increases heterogeneity. In an opposite situation, it could be better to adopt a design that decreases *p,* if it permits one to increase homogeneity. Our results show the importance of investing resighting effort early in the season, i.e. before egg laying, in Black-legged Kittiwakes. They also reveal potential problems with focusing most of the sampling effort at the end of the breeding season (even if at this time the detection probability is high for a good proportion of breeding individuals). Indeed, heterogeneity with respect to resightings can cause bias in survival estimates. Another problem is that heterogeneity can also affect initial captures, such that if captures for marking occur late in the season (as this is often the case in Kittiwakes), the sample of marked birds, instead of being random, will mainly consist in birds that were successful or at least that bred on successful cliffs. In addition, other field work and gathering of informative data may be required for a study (e.g., handling of nestlings for banding, recording of breeding performance, laying/hatching date, etc.) and will condition the attribution of sampling effort over the course of the breeding season. All these aspects should ideally be considered in full design optimization. In relation to our findings, it would for instance be especially interesting to explore how the overall timing of the season (e.g., measured by the average laying date) varies among years and how this affects the pattern of within-year resighting of individuals in a multi-year monitoring program.

Clearly, our results show that, in species like Kittiwakes, a few well-timed sampling occasions can be very efficient for providing reliable data for survival rate estimation. Few details are usually provided in published studies about within-year sampling effort in such species (e.g., Sandvik et al. 2005; but see Golet et al. 2004) and, in these cases, one may question the accuracy of survival estimates and whether changes in the timing of sampling effort might not have improved parameter estimation. In such studies, one could for example expect some interesting improvements by using a multi-state robust design approach (Kendall et al. 2009, Converse et al. 2009). It is likely that the additional effort would be offset by the benefits in terms of estimation accuracy. Our results further show that considering the local reproductive success may be very important because the response of birds may be very different in successful and unsuccessful areas. This difference can have direct conservation implications, notably when monitoring the dynamics of subdivided populations, in which emigration from local study plots can be confounded with mortality. The problems of detectability raised in this study are likely to exist for the other colonial birds and more generally for all species with extensive breeding seasons. Our results indicate that detailed knowledge about the population biology of the species under study can be extremely useful for optimizing sampling design (see also Kendall et al. 2009; Converse et al. 2009).

Acknowledgments We thank R.T. Barrett, R. Garnier and N.G. Yoccoz for assistance at various stages of the work, as well as an anonymous referee for his/her comments. This work benefited from support by the French Polar Institute (IPEV, program no 0333), the Centre National de la Recherche Scientifique (CNRS) and the Agence National de la Recherche (ANR). We also thank Fylkesmannen i Finnmark and Kystverket (Norway) for allowing us to carry out this work on Hornøya.

References

Anker-Nilssen T, Bakken V, Strøm H, Golovkin AN, Bianki VV, Tatarinkova IP (eds) (2000) The status of marine birds breeding in the Barents sea region. Norsk Polarinstitutt, Tromsø

Barrett RT (2001) Monitoring the Atlantic puffin *Fratercula arctica*, common guillemot *Uria aalge* and black-legged kittiwake *Rissa tridactyla* breeding populations on Hornøya, northeast Norway, 1980–2000. Fauna Norvegica 21:1–10

Boulinier T, Danchin E, Monnat JY, Doutrelant C, Cadiou B (1996) Timing of prospecting and the value of information in a colonial breeding bird. J Avian Biol 27:252–256

Boulinier T, Yoccoz NG, McCoy KD, Erikstad KE, Tveraa T (2002) Testing the effect of conspecific reproductive success on dispersal and recruitment decisions in a colonial bird: design issues. J Appl Stat 29:509–520

Boulinier T, McCoy KD, Yoccoz NG, Gasparini J, Tveraa T (2008) Public information affects breeding dispersal in a colonial bird: kittiwakes cue on neighbours. Biol Lett 4:538–540

Buckland ST (1982) A Mark-Recapture survival analysis. J Anim Ecol 51:833–847

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York

Burnham KP, Anderson DR, White GC, Brownie C, Pollock KP (1987) Design and analysis of methods for fish survival experiments based on release–recapture. American Fisheries Society Monograph 5, Bethesda, Maryland

Carothers AD (1973) Effects of unequal catchability on Jolly-Seber estimates. Biometrics 29:79–100

Carothers AD (1979) Quantifying unequal catchability and its effect on survival estimates in an actual population. J Anim Ecol 48:863–869

Cézilly F, Viallefont A, Boy V, Johnson AR (1996) Annual variation in survival and breeding probability in greater flamingos. Ecology 77:1143–1150

Choquet R, Lebreton J-D, Gimenez O, Reboulet A-M, Pradel R (2009a) U-CARE: Utilities for performing goodness of fit tests and manipulating capture-recapture data. Ecography 32:1071–1074

Choquet R, Rouan L, Pradel R (2009b) Program E-SURGE: a software application for fitting mulitevent models. Environ Ecol Stat 3:845–866

Clobert J (1995) Capture recapture and evolutionary ecology: a difficult wedding? J Appl Stat 22:989–1008

Clobert J, Julliard R, McCleery RH (1993) The components of local recruitment. In: Lebreton J-D, North PM (eds) Marked individuals in the study of bird populations. Birkhauser, Basel, pp 281–293

Converse SJ, Kendall WL, Doherty PF Jr, Naughton MB, Hines JE (2009) A traditional and less-invasive robust-design: choice in optimizing effort allocation for seabird population studies. Environ Ecol Stat Ser 3:727–744

Crespin L, Harris MP, Lebreton J-D, Wanless S (2006) Increased adult mortality and reduced breeding success with age in a population of common guillemot *Uria aalge* using marked birds of unknown age. J Avian Biol 37:273–282

Crespin L, Choquet R, Lima M, Merritt J, Pradel R (2008) Is heterogeneity of catchability in capture-recapture studies a mere sampling artifact or a biologically relevant feature of the population? Popul Ecol 50:247–256

Danchin E, Boulinier T, Massot M (1998) Habitat selection based on conspecific reproductive success: implications for the evolution of coloniality. Ecology 79:2415–2428

Davis SA, Akison LK, Farroway LN, Singleton GR, Leslie KE (2003) Abundance estimators and truth: accounting for individual heterogeneity in wild house mice. J Wildl Manage 67:634–645

Drickamer LC, Feldhamer GA, Mikesic DG, Holmes CM (1999) Trap-response heterogeneity of house mice (*Mus musculus*) in outdoor enclosures. J Mammal 80:410–420

Faustino CR, Jennelle CS, Connolly V, Davis AK, Swarthout EC, Dhondt AA, Cooch EG (2004) *Mycoplasma gallisepticum* infection dynamics in a house finch population: seasonal variation in survival, encounter and transmission rate. J Anim Ecol 73:651–669

Frederiksen M, Harris MP, Wanless S (2005) Inter-population variation in demographic parameters: a neglected subject? Oikos 111:209–214

Gilbert RO (1973) Approximations of the bias in the Jolly-Seber capture-recapture model. Biometrics 29:501–526

Golet GH, Schmutz JA, Irons DB, Estes JA (2004) Determinants of reproductive costs in the long-lived black-legged kittiwake: a multiyear experiment. Ecol Monogr 74:353–372

Grosbois V, Thompson PM (2005) North Atlantic climate variation influences survival in adult fulmars. Oikos 109:273–290

Hurvich CM, Tsai CL (1989) Regression and time-series model selection in small samples. Biometrika 76:297–307

Hwang W-D, Chao A (1995) Quantifying the effects of unequal catchabilities on Jolly-Seber estimators via sample coverage. Biometrics 51:128–141

Kendall WL, Converse SJ, Doherty PF, Naughton MB, Anders A, Hines JE, Flint E (2009) Sampling design considerations for demographic studies: a case of colonial seabirds. Ecol Appl 19:55–68

Lebreton J-D, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals—a unified approach with case-studies. Ecol Monogr 62:67–118

Nichols JD, Kendall WL (1995) The use of multi-state capture-recapture models to address questions in evolutionary ecology. J Appl Stat 22:835–846

Ogutu JO, Piepho H-P, Dublin HT, Reid RS, Bhola N (2006) Application of mark-recapture methods to lions: satisfying assumptions by using covariates to explain heterogeneity. J Zool 269:161–174

Pledger S, Pollock KH, Norris JL (2003) Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. Biometrics 59:786–794

Pollock KH, Nichols JD, Brownie C, Hines JE (1990) Statistical-inference for capture-recapture experiments. Wildl Monogr 107:1–97

Pollock KH, Marsh HD, Lawler IR, Allredge MW (2006) Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs. J Wildl Manage 70:255–262

Pradel R, Hines JE, Lebreton J-D, Nichols JD (1997) Capture-recapture survival models taking account of transients. Biometrics 53:60–72

Prévot-Julliard AC, Lebreton J-D, Pradel R (1998) Re-evaluation of adult survival of black-headed gulls (*Larus ridibundus*) in presence of recapture heterogeneity. Auk 115:85–95

Sandvik H, Erikstad KE, Barrett RT, Yoccoz NG (2005) The effect of climate on adult survival in five species of North Atlantic seabirds. J Anim Ecol 74:817–831

Sanz-Aguilar A, Tavecchia G, Minguez E, Massa B, Lo Valvo F, Ballesteros GA, Barbera GG, Amengual JF, Rodriguez A, McMinn M, Oro D (2010) Recapture processes and biological inference in monitoring burrow-nesting seabirds. J Ornithol 151:133–146

Senar JC, Conroy MJ (2004) Multi–state analysis of the impacts of avian pox on a population of serins (*Serinus serinus*): the importance of estimating recapture rates. Anim Biodivers Conserv 27:133–146

Summerlin CT, Wolfe JL (1973) Social influences on trap response of the cotton rat, *Sigmodon hispidus*. Ecology 54:1156–1159

Tavecchia G, Pradel R, Lebreton J-D, Biddau L, Mingozzi T (2002) Sex-biased survival and breeding dispersal probability in a patchy population of the rock sparrow *Petronia petronia*. Ibis 144:E79–E87

Viallefont A, Cooke F, Lebreton J-D (1995) Age-specific costs of first-time breeding. Auk 112:67–76

White GC, Burnham KP (1999) Program MARK: survival estimation from populations of marked animals. Bird Study 46:120–139

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

Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. Trends Ecol Evol 16:446–453