27 May 2018

Nick P. Gondek

Conservation Metrics Inc.

145 McAllister Way

Santa Cruz, CA 95060

(414) 581-4445

nick@conservationmetrics.com

**Too much DNA? Subsampling strategies for Spatially Explicit Capture-Recapture Estimators**

NICK P. GONDEK,[[1]](#footnote-1) *University of Minnesota, 2003 Upper Buford Cir, St Paul, MN 55108 USA*

DAVID L. GARSHELIS, *Minnesota Department of Natural Resources, 1201 E. Hwy 2, Grand Rapids, MN 55744, USA*

KAREN V. NOYCE, *Minnesota Department of Natural Resources, 1201 E. Hwy 2, Grand Rapids, MN 55744, USA, now retired*

JOHN F. FIEBERG,[[2]](#footnote-2) *University of Minnesota, 2003 Upper Buford Cir, St Paul, MN 55108 USA*

**ABSTRACT** Genetic mark-recapture studies estimate animal abundance using non-invasive DNA identification methods to "capture" and subsequently "recapture" individuals that leave genetic material at trap sites. Due to the cost of genotypic analysis, researchers often choose to process only a subsample of this genetic material. We explored the effect of subsampling on spatially-explicit capture recapture (SECR) estimators using hair-snare data obtained from a 2012 genetic mark-recapture study of black bears (*Ursus americanus*), and by simulating capture histories from a known population. Similar to effects on non-spatial mark-recapture estimators, subsampling produced density estimates that were lower, on average, than the full data estimate when individuals left varying numbers of samples at a trap; however non-proportional subsampling (sampling at least one observation from each unique trapping site and sampling period combination) had much less of an effect on estimator performance than simple random sampling. Thus, non-proportional subsampling may be preferable to simple random sampling, despite the inherent violations of SECR assumptions that may result. The benefits of using non-proportional sampling will be greatest when individuals exhibit substantial heterogeneity in their capture propensities, leave multiple samples at a trap, and when available funding severely limits the number of samples that can be processed.

Wildlife managers routinely use mark-recapture studies to estimate animal abundance; especially in the case of endangered species and game animals, abundance and its associated temporal trends are of critical importance for making informed management decisions (Borchers *et al.* 2002; McCrea and Morgan 2014). Abundance estimates can be difficult to interpret, however, without some understanding of the effective area sampled; extrapolation to a regional scale by way of density is almost always desired, if not necessary (Borchers *et al.* 2002; Royle *et al.* 2013). As such, abundance estimates without associated reliable density estimates may be of limited use to managers.

Ad-hoc methods are sometimes used to estimate the effective area sampled in mark-recapture studies (e.g., by quantifying how far animals move); this area can then be used to rescale abundance estimates to obtain estimates of density (Wilson and Anderson 1985, White and Shenk 2001, Ivan et al. 2013ab). Spatially-explicit capture recapture methods (SECR), by contrast, link abundance estimators to their associated study area in a statistically rigorous way (Efford *et al.* 2005; Borchers 2012; Royle *et al.* 2013). SECR estimates of abundance scale directly with a given sampled area, potentially allowing for better standardization of estimates across space or time. SECR models also make better use of capture-recapture data by using the information in the spatial capture histories to model detection probabilities as a function of the distance between each animal's (latent) activity center (AC) and the trapping grid. As such, SECR provides a way to account for an important source of individual heterogeneity in capture probabilities that is unaccounted for in traditional (non-spatial) mark-recapture estimators (e.g., animals with home ranges near the edge of the trapping grid will be less likely to be captured than animals with activity centers near the center of the trapping grid).

SECR models are often fit to capture histories obtained using genetic data (e.g., from hair collected at a trap site), which provides a non-invasive and efficient means of detecting unique individuals (Boulanger *et al.* 2004; Petit and Valiere 2006; Gervasi *et al.* 2008). Due to the costs of genotypic analysis, however, managers may only be able to process a subsample of genetic material left at trapping sites (Boulanger *et al.* 2004; Petit and Valiere 2006; Gervasi *et al.* 2008; Settlage *et al.* 2008).

Subsampling has been shown to be problematic for non-spatial mark-recapture estimators, especially when individuals exhibit a behavioral response to having been previously captured and this behavioral response is not consistent across individuals (Tredick *et al.* 2007; Ebert *et al.* 2010; Augustine *et al.* 2014). In this case, individuals that leave many DNA samples (e.g., clusters of hair) are likely to be identified in a subsample, whereas individuals that leave few samples are often excluded. In other words, samples of DNA selected in a subsample of data are likely to come from individuals that are repeatedly captured. As a result, subsampling tends to result in estimators of capture probability that are biased high and abundance estimators that are biased low; this effect is more pronounced as subsample size decreases (Augustine *et al.* 2014).

A fundamental difference between SECR and non-spatial capture-recapture models relates to how the estimators make use of multiple captures of the same individual during the same time period. Whereas non-spatial models collapse multiple captures at different traps (same time period) into a single capture event, these multiple captures would be used to inform individual movement characteristics in a SECR model (Borchers 2012; Royle et al. 2013). Thus, samples that are redundant in a non-spatial model (captures at trap location in a single period) are of critical importance to SECR models. Importantly, managers often avoid taking multiple samples from a single trap under the assumption that they are likely to be from the same individual, and thus not informative and not worth the cost of genotypic analysis. This strategy is effective for use with non-spatial models; analyzing several samples from the same site often leads to diminishing returns in precision and accuracy (Dreher *et al.* 2009). Instead, preference is given to samples from novel site by session combinations (hereafter referred to as site-sessions), because they are more likely to represent new individuals. Yet, this approach tends to increase inclusion probabilities for samples left at infrequently visited sites and decrease inclusion probabilities for samples left at frequently visited sites. Because SECR models attempt to use information in the spatial capture histories to inform capture probabilities (by modeling movement about activity centers), we hypothesized that the effects of non-proportional subsampling would be problematic for SECR models (more so than for non-spatial capture-recapture models that reduce capture histories to a single binary response for each sampling period). A natural solution to this violation is to avoid preference for novel site-sessions (i.e., use simple random sampling), at the cost of identifying fewer unique individuals.

Others have shown that subsampling results in non-spatial mark-recapture estimators that are biased low, but the magnitude of the bias depended on the data set and several assumptions such as erroneous genotyping rate and trap spacing (Tredick *et al.* 2007; Dreher *et al.* 2009; Augustine *et al.* 2014, Murphy *et al*. 2016). The effects of subsampling using different sampling methods has rarely been investigated in the context of SECR models (but see the Appendix A of Murphy *et. al.* 2016). Our objective was to use a northern Minnesota genetic mark-recapture data set in tandem with simulated data sets with known individuals, to compare abundance and density estimates using (1) subsampling strategies commonly utilized by wildlife managers and researchers and (2) various subsampling rates reflective of different budgetary constraints. Using these results, we provide guidance for obtaining genetic SECR density estimates when budget constraints preclude the possibility of processing all samples of genetic material.

**STUDY AREA**

**METHODS**

## Empirical data set

We used empirical data from a 2012 genetic mark-recapture study of American black bears (*Ursus americanus*) in northern Minnesota (Garshelis and Noyce 2013). For this study, we constructed 121 hair-snare traps across a grid of […] km2 cells. We used 2-strand barbed wire corral-type traps ( ), baited in the center with inaccessible suspended bacon and scent lure plus a small food reward on the ground. We checked every site once during each of six, 10-day trapping sessions from May through July 2012. We collected all hair snagged during each session, comprising 2784 total samples and 1642 separate clusters of samples (hairs collected from adjacent barbs and assumed to be from the same animal). We submitted one sample each from 1113 of these clusters for genotypic analysis (Wildlife Genetics International, Nelson, British Columbia); hair samples from 1019 clusters were successfully linked to individual bears. We set camera traps at a subset of the hair trap locations to provide information on how common it was for individual bears to visit the same trap multiple times in a single session and to use multiple locations to enter or exit traps.

## Simulated Data

Investigated Scenarios — We simulated eight scenarios with different combinations of behavior, individual heterogeneity, sample redundancy, and uneven density of activity centers (Fig. 4). In all scenarios, we simulated capture histories for 30 bears during 6 time periods on a 6 by 6 trapping grid with traps spaced 800 meters apart, for a total non-buffered area of 23.04 km2. The number of bears and size of the trapping grid were chosen to roughly mimic the empirical data set up. Rather than basing our simulations on a desired number of DNA samples for each simulation, we instead modified the likelihood of capture, recapture, and propensity of bears to leave redundant samples in each scenario, allowing us to quantify impacts of subsampling on estimator performance across a range of realistic conditions.

*Capture History Simulation. —* We developed a simulation model to create capture histories allowing for the possibility of: 1) capture heterogeneity among individuals; 2) behavioral response to traps (i.e., enhanced attraction to traps following initial capture); and 3) individual bears leaving multiple hair clusters at a site during a single trapping period. To describe the different scenarios, we have compiled a list of terms and definitions used in model notation (Table 1; Fig. S1).

We simulated ‘activity centers’, (*Ai*; *i* = 1, 2, …, *N*), for *N* = 30 individuals using a simple sequential inhibition (SSI) process with an inhibition distance *ϕ* = 200m (Baddeley, 2017).

A ~ rSSI(N, ϕ) (1)

In a simple sequential inhibition process, points are generated at random in the window of interest, and if a new point is generated within *ϕ* distance of an existing point, that point is discarded and a new one is generated. In scenarios with uneven density of activity centers across the trapping grid, we simulated separate SSI processes in two spatial strata, with *Nα* of activity centers located exclusively in the left half of the trapping grid and the remaining activity centers simulated at random from within the entirety of the trapping grid (Fig. 2A).

We assigned a normally distributed ‘individual heterogeneity’ parameter, δi, to each individual, which characterized that individual’s heightened or depressed propensity for capture relative to the population:

δi ~ Normal(0, Δ) (2)

We determined the capture probability for individual *i*, at trap *k*,during trapping session *t, gi,k,t*, using:

where Ψi,k is the distance between the individual’s activity center and trap k, Ci,k ,t is 1 if the individual i was previously captured at trap k during any previous trapping period (and 0 otherwise), and δi measures the individual’s propensity for capture (Fig. 2B). The general form of the model in eq 3 is referred to as a half-normal detection function; g0 determines the maximum probability of detecting an animal and σ represents the rate at which detection drops off with distance between an individual’s activity center and the trap. Note, it is common to use a parameter b to model a change in capture probabilities across all traps following an initial capture and a parameter bk when modeling a change in capture probabilities that applies only to those traps where the individual has been previously captured (see Model Fitting). Here (eq. 3), we use b rather than bk even though we apply this effect only to those traps where the individual has been previously captured, and we pair this parameter with a set of trap-specific indicator variables, Ci,k ,t.. We feel this specification is more natural since the effect of a previous capture is assumed to be the same at every trap where a bear has been previously caught.

Capture histories were then simulated as Bernouli random variables:

c­i,k,t ~Bernouli(gi,k,t) (4)

If captured, we simulated a number of samples left at the trap, Ii,k,t, using a Poisson distribution:

Ii,k,t ~ Poisson(λ­i ci,k,t) with log(λ­i) = γ+δi (5)

By including δi (above), we assume that bears that have a higher propensity for being captured are also more likely to leave more samples at a trap.

## Subsampling

We chose to explore two alternative strategies for subsampling hair left at hair traps in both the empirical and simulated data sets: 1) *simple random sampling* (SRS), and 2) a subsampling method that gives preference to unique site-sessions, which we will refer to as *site-session preferred* (*SPR*). With simple random sampling, *n* samples were chosen at random from the set of hair clusters pooled across the different sites and trapping periods. Alternatively, with *SPR*, we tried to maximize the number of unique site-sessions represented in the subsample. Let *m* represent the number of unique site-sessions in the full data set. If *m > n*, we randomly choose *n* unique site-sessions, with 1 sample randomly selected from each of these site-sessions (in the survey sampling literature, this would be referred to as a 2-stage cluster sample). When *n* > *m*, we randomly chose one sample at random from each unique site-session and then took a second simple random subsample of size *n – m* from the remaining clusters (from the pooled data) to give a total of *n* samples.

We subsampled the empirical (black bear) data set and simulated data sets (from each of the 8 simulation scenarios) using both subsampling strategies, with subsample sizes of *n* = 250, *n* = 550, and *n* = 850. We fit SECR models to each subsampled data set and the original (i.e., “full”) data set. In the context of this study, an observation (i.e. hair cluster) can be considered ‘redundant’ if it does not contribute a unique (individual x site x session) combination to the capture history of interest; i.e., a sample is redundant if there is another observation of the same individual at the particular site-session. To better understand the performance of the estimators under different types of subsampling, we calculated the percentage of each data set that was not-redundant.

**Model Fitting**

A variety of functions can be used to model how detection probabilities change as a function of distance between an animal's activity center and a trap location. Here, we only consider the half-normal curve(Fig. 2B), which we used to simulate capture histories. For each simulated (and potentially subsampled) data set, we fitted two SECR models to the observed capture histories, a null model (*g0* ~ 1), and a model where an individual’s likelihood of capture at a given trap changed after initial capture at that specific trap (*g0* ~ *bk*). (note again, the *bk*parameter in this model is equivalent to theparameter *b* in eq. 3.) For each subsample of the real black bear data, we fit two additional models: a model where likelihood of capture depended on the trapping period (*g0* ~ *t*), and one where capture probabilities varied by trapping period and depended on whether the animal had been previously caught at the trap (*g0* ~ *bk* + *t*). In each case, we assumed was constant for all individuals.

We fit models using the R programming language (R Core Team 2015), package ‘secr’, and packages ‘foreach’ and ‘doParallel’ for optimization of model fitting and capture history simulation using parallel processing (Analytics and Weston 2014, 2015; Efford 2015). Within the package 'secr', we used the function 'secr.fit' for fitting models to subsampled data. This function requires a capture history and a trapping grid and returns a derived density estimate, along with estimated parameters describing the effect of time, trap-specific behavioral responses, and any other individual-level covariates (e.g., sex) on capture probabilities (Efford *et al.* 2005).

## Simulation Process

Note the distinction between a capture history simulation, which refers to the simulation of a set of capture histories under a specific scenario described in Fig. 3 (see *Capture History Simulation*), and a subsampling simulation, which refers to the subsampling of an empirical or simulated capture history data set and subsequent model fitting. We generated 220 sets of capture histories for each of the eight unique simulated bear populations. We then subsampled each set of capture histories using both SPR and SRS sampling designs with sample size equal *n* = 250, 550, or 850. Thus, each simulated set of capture histories was subsampled 6 times (2 methods x 3 sample sizes). We also subsampled the real black bear capture history data set 220 times using both SPR and SRS sampling designs with *n*  = 250, 550, or 850. We then fit SECR models (*g0* ~1, *g0* ~ *bk*) to the subsampled, simulated capture histories and models (*g0* ~1, *g0* ~ *bk*, *g0* ~ *t*, *g0* ~ *t* + *bk*) to the subsampled real capture histories (see *Model Fitting*), and saved the resultant model objects for latter comparison.

# RESULTS

**Empirical Data set**

It was common for there to be multiple clusters of hair snagged at a site during a single sampling session. Photos from cameras placed at a subsample of traps indicated that bears sometimes visited a site multiple times during a single session and/or used multiple locations to enter or exit the trap, even during the same session. The number of samples left at a given site-session by individual bears thus showed wide variation; in 46.7% of cases, bears left only a single cluster of hair (Garshelis and Noyce 2013), whereas in 25.8% of cases, bears left three or more clusters of hair (up to 11) at a given site-session (Fig. 1).

Density estimates derived from fitting SECR models to subsets of *n* = 250 observations tended to be lower, on average, than estimates derived from the full empirical data set (Fig. 4, Fig. S2). Differences between estimates from subsampled and full data sets were greatest when using SRS (versus SPR), and these differences became smaller as the size of the subsampled data sets increased (Fig S2). Lastly, including a trap response (*bk*) as a covariate in the half-normal detection function resulted in estimates from both SRS and SPR that were more consistent with the estimate from the full data set (Fig. 4, Fig S2).

**Simulated Data sets**

As expected, density estimators were biased when the model was mis-specified. Examples included scenarios where: (1) bears left multiple samples at a trap and a behavioral effect was present but not included in the SECR model (t6, t7), (2) individuals exhibited unmodeled heterogeneity in capture probabilities (t3, t5-t7), or (3) activity centers were not uniformly distributed (t7-t8; Fig. 3, Fig. 5). Similar to the empirical data set, 3 of 8 simulated capture scenarios (t5, t6, and t7) yielded SECR density estimates that tended to be lower, on average, than those derived from the full data set, and for these 3 scenarios, estimates from SRS samples were lower than those from SPR samples (Fig 4). In each of these scenarios, bears exhibited individual heterogeneity in their capture propensities and also left multiple samples at a trap. SPR estimates were also less variable than those obtained from SRS samples in these scenarios and in Scenario t4, where bears deposited multiple samples at a trap but there was no unmodeled heterogeneity (Fig. 4). Further, SPR subsampling resulted in lower proportions of redundant samples, on average, than SRS subsampling, particularly at lower sample sizes (Fig. 6). SPR and SRS performed similarly in simulations in which bears only deposited 1 sample at a trap (t1, t2, t3 and t8) and at higher sample sizes (Fis. S3). In all cases where models were properly parameterized, subsampled data resulted in estimators of capture probabilities (*gi,k,t*) and behavioral effects that were biased low (Fig. S4, S5).

# DISCUSSION

Though it may seem counter-intuitive, we found that non-random sampling (SPR) outperformed simple random sampling (SRS) when using SECR. For example, estimates of bear density derived using SPR to subsample the data were closer to the true bear density in simulated populations than estimates derived using SRS. They were likewise closer to the estimate derived using the full data set, both in simulated populations and in the empirical study. We recognized that non-proportional sampling could be problematic due to its effect on inclusions probabilities – samples left at infrequently visited sites would be more likely to be included in the subsample than samples left at frequently visited sites. Potentially, these variable inclusion probabilities could result in biased estimators of detection parameters (e.g., *g0* and *σ*) and ultimately, density.

Indeed, we found that both within the constructs of our 8 simulated capture history scenarios and in the empirical black bear data set, any potential bias from SPR was outweighed by the selection of fewer redundant samples than SRS. The benefits of using non-proportional sampling (SPR) were greatest when individuals exhibited substantial heterogeneity in their capture propensities and left multiple samples at a trap, both of which were clearly true in the empirical study and are widely acknowledged to be true by bear researchers conducting hair-trapping studies ( ). The reduced bias of SPR sampling relative to SRS sampling was most evident when relatively few samples were processed, which is often the case in field studies with constrained budgets.

Our results mirror those found in studies investigating non-spatial mark-recapture estimators with missing data resulting from subsampling or failure to genotype, in that post-sampling behavioral effects were biased low (Tredick et. al 2007, Augustine et. al 2014). These biases were present whenever bears left multiple samples at the trap (t4-t7), and they were most notable when individual heterogeneity and redundancy were both present (t5-t7).

Bears in the empirical study exhibited substantial heterogeneity in the number of samples deposited at a given site-session; in 47% of cases, bears left only a single sample at a given site-session, but some left as many as 11 (Fig. 1, Noyce and Garshelis 2013). Thus, subsampling empirical data using SRS was highly likely to select redundant data from bears that tended to leave several samples at a single site-session. Conversely, SPR performed well because much of the data it excluded from the full data set was redundant (repeated individual by site by session combinations).

Both SECR and non-spatial mark-recapture estimators are biased when unmodeled heterogeneity in capture probabilities exists within the study population. Similarly, traditional SECR estimators are biased when activity centers are not uniformly distributed (as in scenario t7 and t8). We note that it is possible to model individual heterogeneity using finite mixture models (Borchers and Efford 2008), and to model spatial variation in the density of activity centers using habitat covariates (Royle *et. al* 2013). Recently developed categorical spatial partial identity models (Categorical SPIM) also allow the use of partially identified genetic samples, which are often excluded from analysis in the SECR framework due to the “shadow effect” - erroneously treating novel individuals as recaptures due to having similar genotypes (Mills *et al.* 2000, Augustine *et al.* 2018). We suspect it may be possible to develop SECR models that accommodate non-SRS subsampling designs. For the scenarios we considered, however, the effects of subsampling on SECR density estimates were relatively minor.

**MANAGEMENT IMPLICATIONS**

Genetic mark-recapture studies frequently result in more DNA samples than researchers can afford to process. In these cases, it is best to choose samples to process using a strategy that maximizes the number of unique site-sessions in the processed data set. We proposed to randomly choose 1 sample from unique site-sessions, and then choose any additional samples using simple random sampling (i.e., a random selection from the remaining pooled data). This strategy resulted in density estimates that were less variable and more accurate than estimates obtained using simple random sampling, particularly when animals displayed individual heterogeneity in their propensity for capture. The benefits of using this subsampling approach are expected to increase as subsample size decreases, and particularly when individuals leave multiple DNA samples at a trap.

# ACKNOWLEDGEMENTS

# This project was funded in part by the Minnesota Department of Natural Resources (Game and Fish Fund) and Federal Wildlife Restoration (Pittman-Robertson) Program. JF received partial support from the Minnesota Agricultural Experimental Station and the McKnight Foundation.

# LITERATURE CITED

Analytics, R., and S. Weston. 2014. doParallel: Foreach parallel adaptor for the parallel package. <http://cran.r-project.org/package=doParallel>. Accessed 15 November, 2015.Analytics, R., and S. Weston. 2015. foreach: Provides Foreach Looping Construct for R. <http://cran.r-project.org/package=foreach>. Accessed 15 November, 2015.

Augustine, B. C., C. A. Tredick, and S. J. Bonner. 2014. Accounting for behavioural response to capture when estimating population size from hair snare studies with missing data. Methods in Ecology and Evolution 5:1154–1161.

Augustine, B. C., J. A. Royle, S. Murphy, B. Chandler, R. & J. Cox, J. & K. Marcella. 2018. Spatial Capture-Recapture for Categorically Marked Populations with An Application to Genetic Capture-Recapture. Preprint, BioRxiv 10.1101/265678.

Baddeley, A. 2017. spatstat: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests. <http://cran.r-project.org/package=spatstat>. Accessed 15 November, 2015.

Borchers, D. L. and Efford, M. G. 2008. Spatially explicit maximum likelihood methods for capture–recapture studies. Biometrics 64, 377–385.

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: closed populations. Volume 13. Springer Science and Business Media.

Boulanger, J., S. Himmer, and C. Swan. 2004. Monitoring of grizzly bear population trends and demography using DNA mark-recapture methods in the Owikeno Lake area of British Columbia. Canadian Journal of Zoology 82:1267–1277.

Dreher, B. P., G. J. M. Rosa, P. M. Lukacs, K. T. Scribner, and S. R. Winterstein. 2009. Subsampling Hair Samples Affects Accuracy and Precision of DNA-Based Population Estimates. The Journal of Wildlife Management 73:1184–1188.

Ebert, C., F. Knauer, I. Storch, and U. Hohmann. 2010. Individual heterogeneity as a pitfall in population estimates based on non-invasive genetic sampling: a review and recommendations. Wildlife Biology 16:225–240.

Efford, M. 2015. secr: Spatially explicit capture-recapture models. <http://cran.r-project.org/package=secr>. Accessed 15 November, 2015.

Efford, M. G., B. Warburton, M. C. Coleman, and R. J. Barker. 2005. A field test of two methods for density estimation. Wildlife Society Bulletin 33:731–738.

Garshelis, D. L., and K. V Noyce. 2013. Capture heterogeneity in hair-trapping of bears. Pages 71–85 in L. Cornicelli, M. Carstensen, M. Grund, M. Larson, and J. SW. Lawrence editors. Summaries of wildlife research findings 2015. Minnesota Department of Natural Resources. St. Paul, USA.

Gervasi, V., P. Ciucci, J. Boulanger, M. Posillico, C. Sulli, S. Focardi, E. Randi, and L. Boitani. 2008. A preliminary estimate of the Apennine brown bear population size based on hair-snag sampling and multiple data source mark-recapture Huggins models. Ursus 19:105–121.

Ivan, J. S., G. C. White, and T. M. Shenk. 2013a. Using auxiliary telemetry information to estimate animal density from capture–recapture data. Ecology 94:809–816.

Ivan, J.S., White, G.C. and Shenk, T.M., 2013b. Using auxiliary telemetry information to estimate animal density from capture–recapture data. Ecology, 94(4), pp.809-816.

Laake, and J.L. 2013. {RMark}: An R Interface for Analysis of Capture-Recapture Data with {MARK}. Seattle, {WA}. <http://www.afsc.noaa.gov/Publications/ProcRpt/PR2013-01.pdf>.

McCrea, R. S., and B. J. T. Morgan. 2014. Analysis of Capture-Recapture Data. Chapman and Hall, CRC.

Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, and D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. Ecological applications 10:283–294.

Murphy, S. M., J. J. Cox, B. C. Augustine, J. T. Hast, J. M. Guthrie, J. Wright, J. McDermott, S. C. Maehr, and J. H. Plaxico. 2016. Characterizing recolonization by a reintroduced bear population using genetic spatial capture–recapture. Journal of Wildlife Management 80:1390–1407.

Petit, E., and N. Valiere. 2006. Estimating population size with noninvasive capture-mark-recapture data. Conservation Biology 20:1062–1073.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria. <http://www.r-project.org/>. Accessed 15 November, 2015.

Royle, J. A., R. B. Chandler, R. Sollmann, and B. Gardner. 2013. Spatial capture-recapture. Academic Press.

Settlage, K. E., F. T. MANEN, J. D. Clark, and T. L. King. 2008. Challenges of DNA-Based Mark-Recapture Studies of American Black Bears. The Journal of Wildlife Management 72:1035–1042.

Tredick, C. A., M. R. Vaughan, D. F. Stauffer, S. L. Simek, and T. Eason. 2007. Sub-sampling genetic data to estimate black bear population size: a case study. Ursus 18:179–188.

Wilson, K. R., and D. R. Anderson. 1985. Evaluation of two density estimators of small mammal population size. Journal of Mammalogy 66:13–21.

White, G. C., and T. M. Shenk. 2001. Population estimation with radio-marked animals. Pages 329–350 in J. J. Millspaugh and J. M. Marzluff, editors. Radio tracking and animal populations. Academic Press, San Diego, California, USA.

Figure 1. Histogram displaying the number of samples deposited across unique site-sessions by individual black bears (Ursus americanus) in a 2014 genetic mark-recapture study in northern Minnesota.

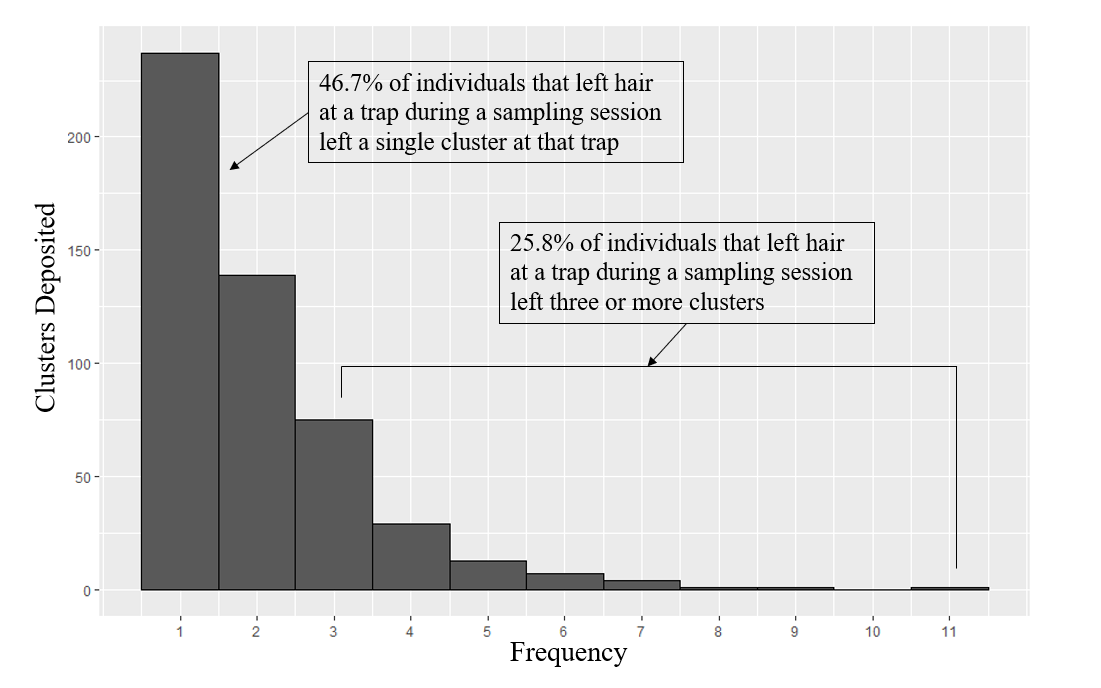
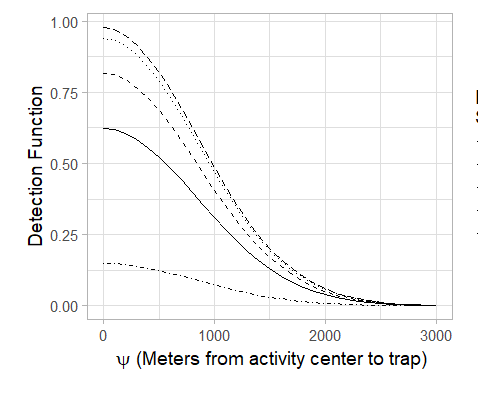
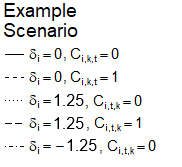
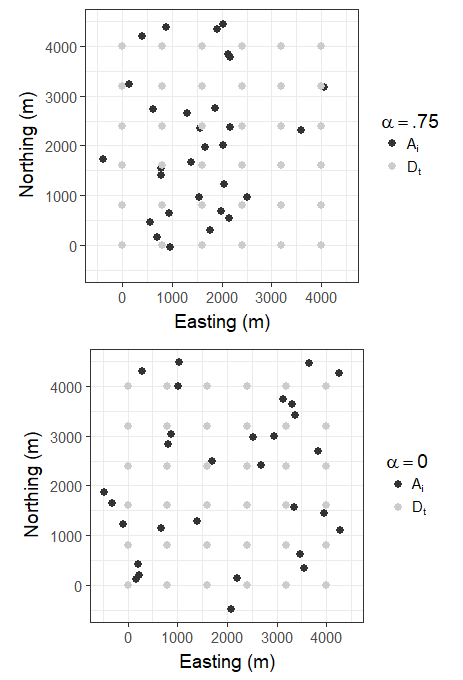


Figure 2. A) Example “Activity Centers” (black) generated using either a heavy skew towards activity centers being located in one half of the grid (α = 0.75) or an absence of skew in the location of activity centers (α=0). Trap locations are represented as gray dots. (B) Example of half-normal capture probability curves for individuals*,* dependent on their individual propensity of capture, whether the given trap has captured the individual in a previous session ( if previously captured, 0 if not), and their distance from the trap (d), as defined by Equation 3, . Capture probabilities are higher when an individual has been previously captured at a trap ( and for individuals with large heterogeneity parameters (.

Figure 3. Matrix of behavioral and density effects included in 8 scenarios used to generate the 8 simulated capture histories used in this study. Shaded boxes represent the presence of a given effect in the simulated capture histories.

B

A

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **Description of effect** | **Effect scale** | **Present in** | | | | | | | | **Terms** | |
| **t1** | **t2** | **t3** | **t4** | **t5** | **t6** | **t7** | **t8** | **If Absent** | **If Present** |
| Positive trap-specific behavior | Bears are more likely to revisit a given trap after visiting that trap | Population |  |  |  |  |  |  |  |  | b = 0 | b = 1 |
| Individual behavioral heterogeneity | Individual bears are more or less likely to visit any trap | Individual |  |  |  |  |  |  |  |  | Δ = 0,  δ = 0 | Δ = 1.25,  δ ~ N(0, Δ) |
| Sample redundancy | Bears leave >1 sample at a site-session | Population if Δ = 0; Individual if Δ > 0 |  |  |  |  |  |  |  |  | λi = 0 | λi = e(γ + δi) |
| Uneven density of activity centers | Bears are distributed disproportionately on trapping grid | Population |  |  |  |  |  |  |  |  | α = 0 | α = .75 |

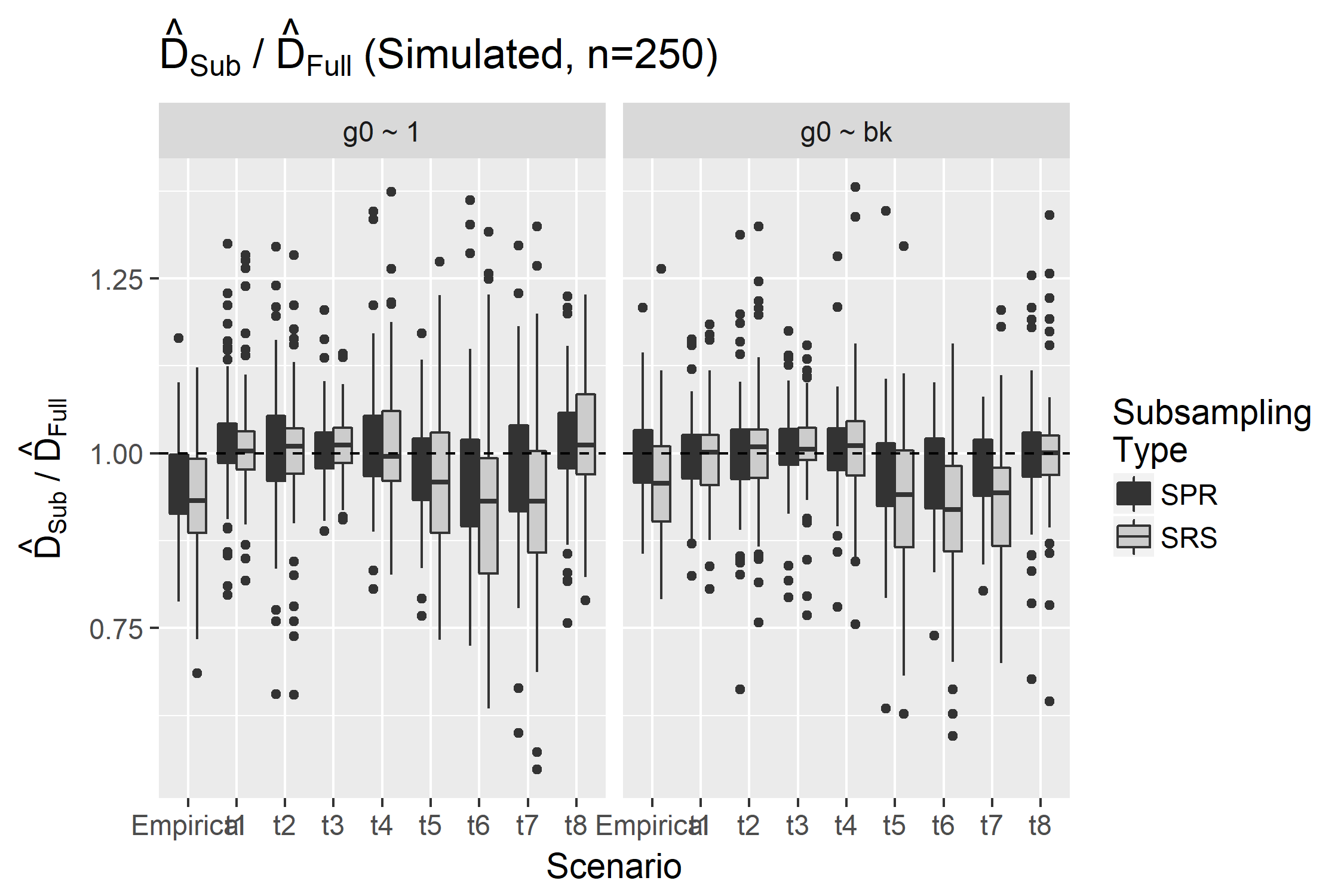
Figure 4. Ratio of density estimates obtained using subsamples of the simulated and empirical data sets (n = 250) relative to the estimates obtained by fitting the given model on the full data set (. Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. Simulation scenarios incorporated trap specific behavioral effects (t2, t6, t7), individual heterogeneity in capture probability (t3, t6, t7), redundancy in sample deposition (t4, t5, t6, t7), and/or uneven distribution of activity centers (t7 and t8; Figure 4) with 30 individuals over 6 trapping periods. Empirical data (labelled ‘Empirical’) were collected from individual black bears (Ursus americanus) from May through July 2012 in a genetic mark-recapture study in northern Minnesota. Data were subsampled using either Simple Random Sampling (SRS) or using an approach that gave preference to unique site-sessions, Site-Session Preferred (SPR), and fitted to both a null model () and a model with a trap-specific behavior covariate (). 

Figure 5. Ratio of density estimates obtained using subsampled simulated data sets ( and full data sets relative to the true density of the simulated population (D), using simulation scenarios incorporating a positive trap specific behavioral effect (t2, t6, t7), individual heterogeneity in capture probability (t3, t6, t7), redundancy in sample deposition (t4, t5, t6, t7), and/or uneven distribution of activity centers (t7 and t8; Figure 4). Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. All simulations included 30 individuals over 6 trapping periods. Data were subsampled using either Simple Random Sampling (SRS) or using an approach that gave preference to unique site-sessions, Site-Session Preferred (SPR), and fitted to both a null model () and a model with a trap-specific behavior covariate ().

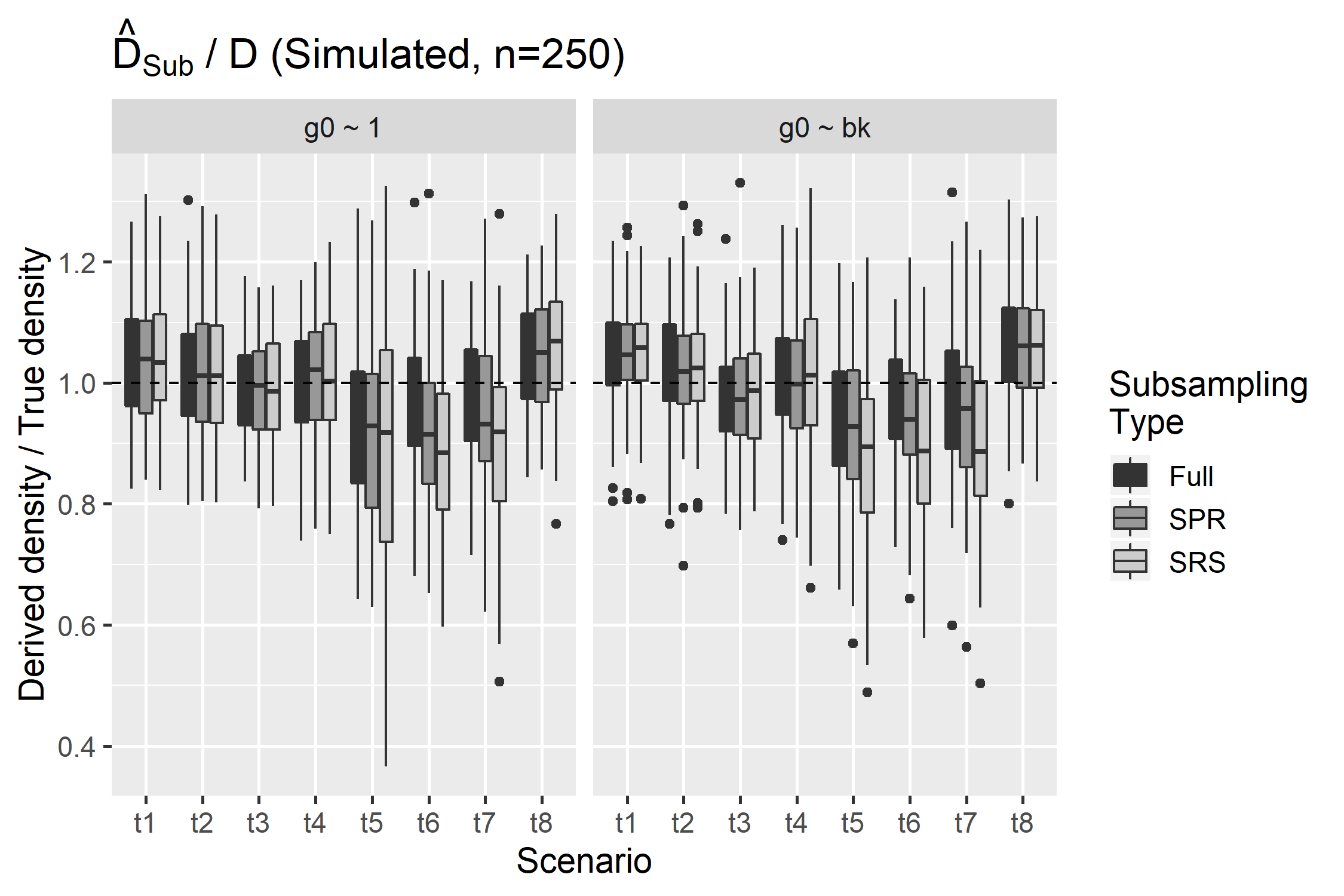


Figure 6. Proportion of non-redundant samples (contribute novel individual by site-session combinations to the capture history) vs subsampling type for each of the four simulated scenarios where redundancy is possible (t4, t5, t6 and t7; Fig 4). Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. Simulation scenarios incorporated a positive trap-specific behavioral effect (t6, t7), individual heterogeneity in capture probability (t6, t7), redundancy in sample deposition (t4, t5, t6, t7), and/or uneven distribution of activity centers (t7; Figure 4) with 30 individuals over 6 trapping periods. Note that, as redundancy is not introduced for scenarios t1, t2, t3, or t8, the proportion of non-redundant samples is fixed at 1 for these scenarios. Data were subsampled using either Simple Random Sampling (SRS) or using an approach that gave preference to unique site-sessions, Site-Session Preferred (SPR).

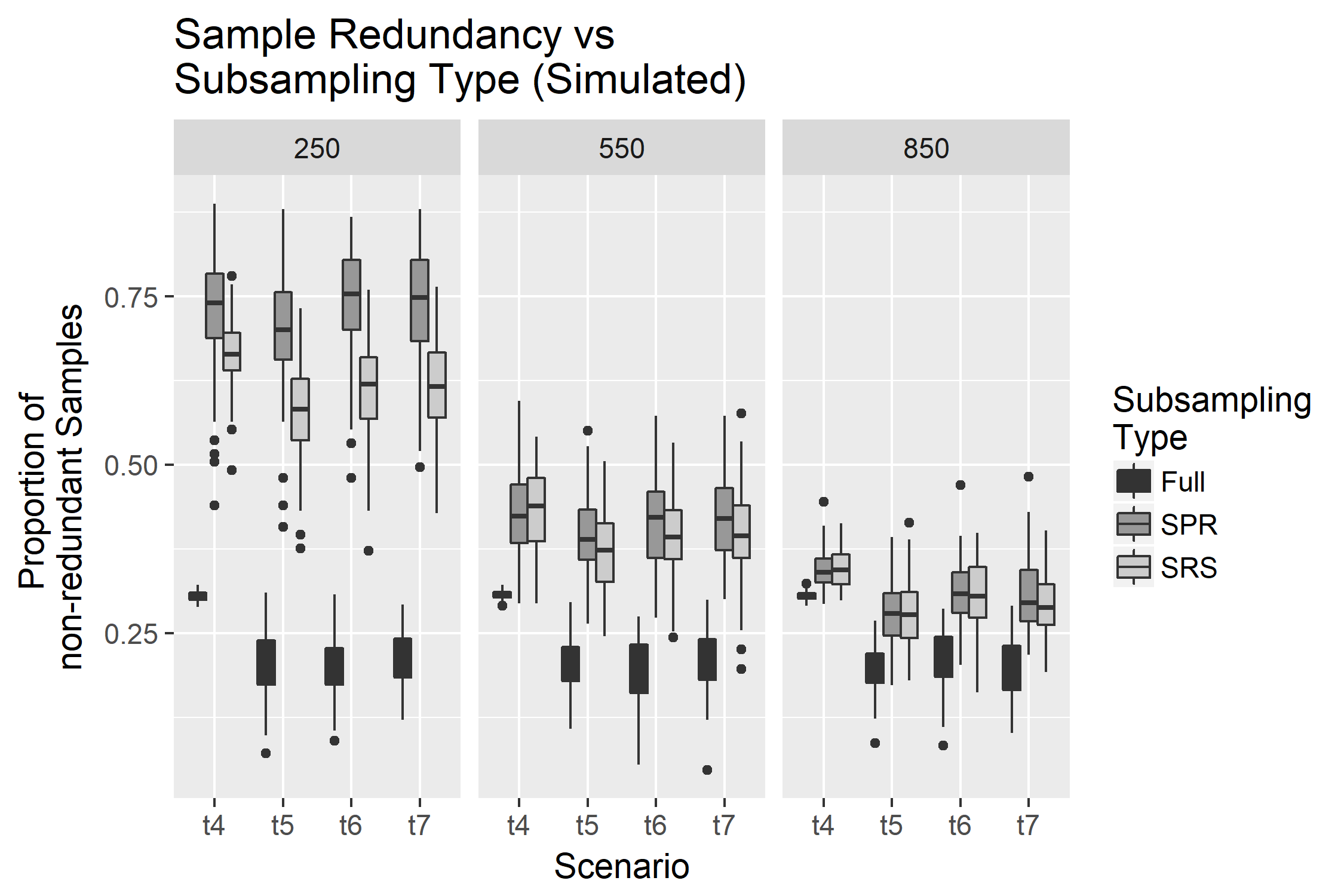


Table 1**.** Model Notation for simulation of capture histories (Eq 1-5)

|  |  |
| --- | --- |
| Term | Definition |
| *N* | Size of bear population (30 in all scenarios) |
| *K* | Number of traps on trapping grid during a single simulation (36 in all scenarios) |
| *Dk* | Location of trap *k* on trapping grid during a single simulation{*k* = 1, ..., *K*} |
| *T* | Number of trapping sessions (6 in all scenarios) |
| *Φ* | Inhibition distance between bear activity centers (200m in all scenarios) |
| *α* | Parameter describing intensity of stratification of bear activity centers (.75 in t6 and t7, 0 in all other scenarios) |
| *Ai* | Locations of bear activity centers {*i* = 1, …, *N*} |
| *Ψi,k* | Euclidean distances between the activity center for individual *i*, *Ai*, and the location of trap *k*, *Dk* |
| *g0* | Logit capture probability at a given trap for a bear whose activity center is exactly at that trap (.5 in all scenarios) |
| *b* | Difference between logit capture and recapture probabilities (1 in t2, t5, t6 and 0 in all other scenarios) |
| *σ* | Inflection point of half-normal distribution which describes capture probability as a function of *Ψ* (846m in all scenarios) |
| *δi* | Parameter describing heterogeneity in individual bears’ capture probabilities; this parameter also influences the expected number of redundant samples deposited by an individual. {*i* = 1, …, *N*} |
| *Δ* | Standard deviation of normal distribution of *δi* values (1.25 in t3, t5, t6 and t7, 0 in all other scenarios) |
| *γ* | Parameter describing log expected number of samples deposited by an individual bear after being captured. |
| *gi,t,k* | Capture probability for individual i at trap k during time period t{i = 1,…, N},{k = 1,..., K},{t = 1,..., T} |
| *ci,t,k* | Indicator variable equal to 1 if bear i was captured at trap k during time period t, and 0 otherwise. |
| *Ci,t,k* | Indicator variable equal to 1 if bear i was captured at trap k at any time before period t, and 0 otherwise {i = 1, …, N},{k = 1, ..., K},{t = 1,..., T} |
| *Ii,t,k* | Number of samples deposited and collected during a simulation {i = 1, …, N},{k = 1, ..., K},{t = 1, ..., T} |

**APPENDIX A**

Figure S1. Directed Acrylic Graph of the data simulation process. Triangle nodes represent fixed parameters, circles represent stochastic values, and squares represent deterministic values obtained using stochastic values. All values are defined in Table 1. Ii,k,t is either a deterministic or stochastic node depending on the value of γ.

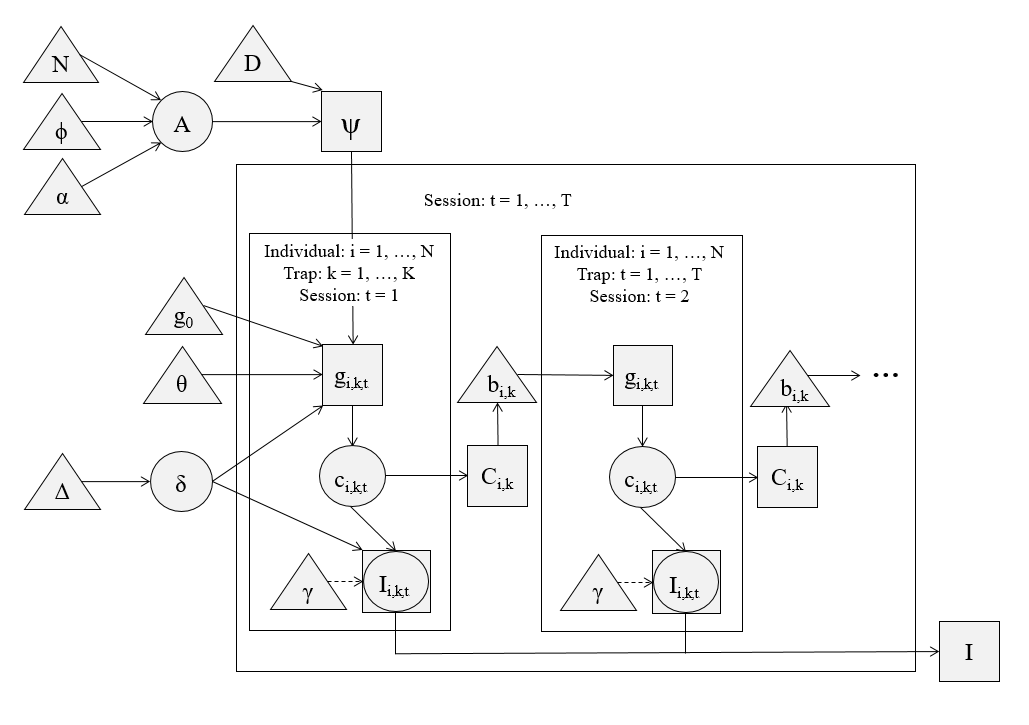


Figure S2: Ratio of density estimates obtained using subsamples of the empirical data (Sub; n = 250, 550 and 850) relative to the estimates obtained by fitting the given model on the full empirical data set (Full) for all scenario and sample size combinations explored.Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. Empirical data were collected from individual black bears (Ursus americanus) from May through July 2012 in a genetic mark-recapture study in northern Minnesota.

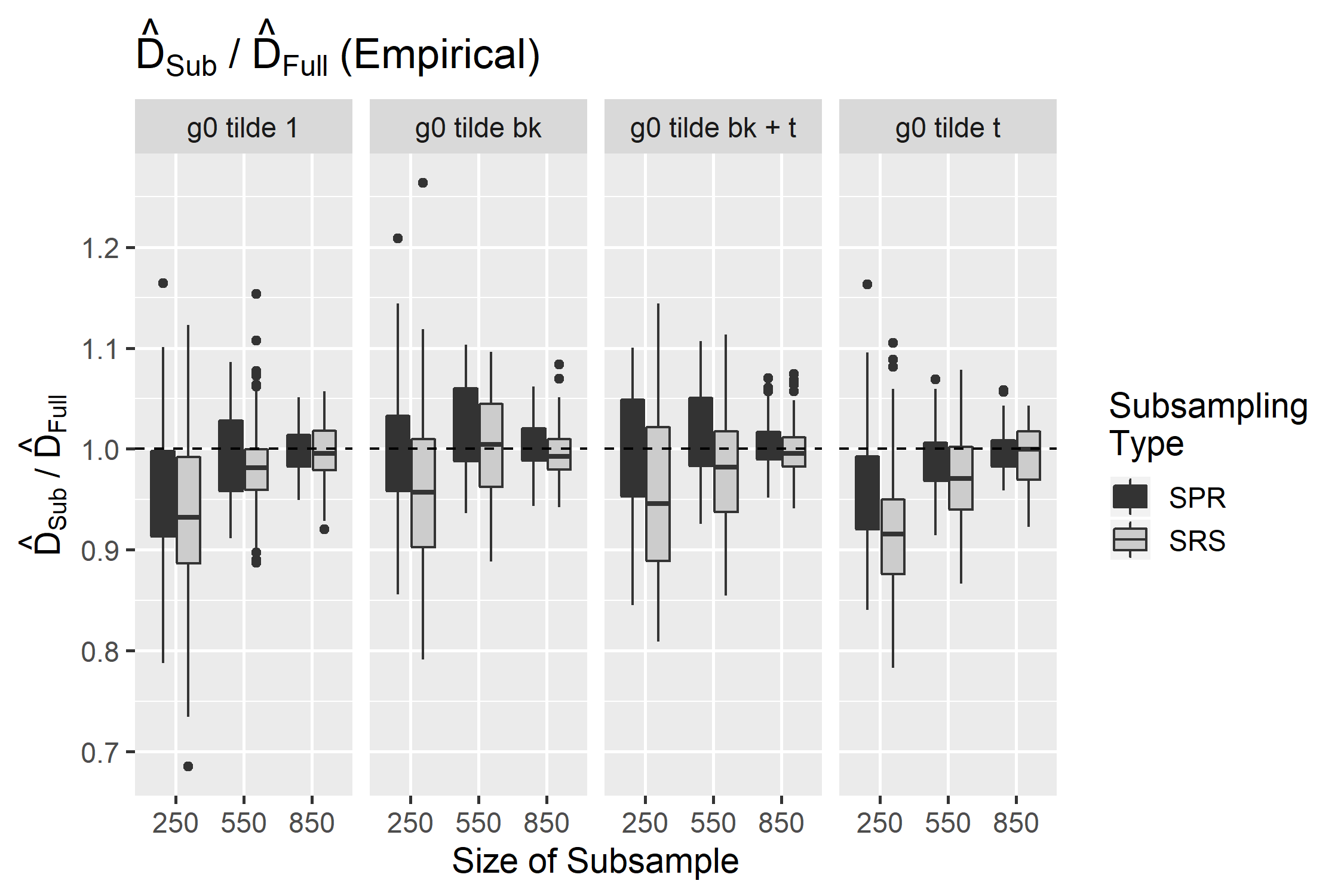


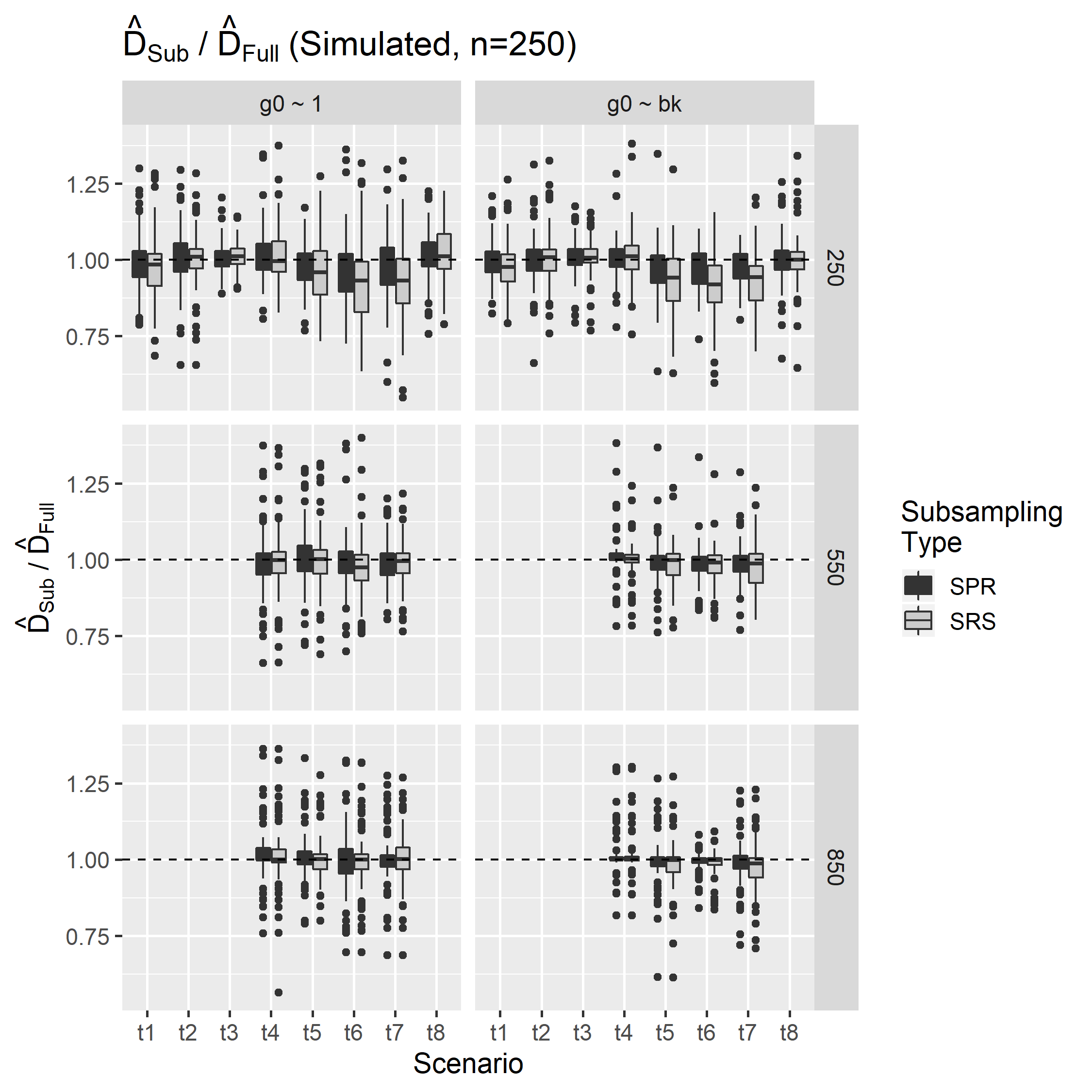
Figure S3: Ratio of density estimates obtained using subsamples of the simulated data (Sub; n = 250, 550 and 850) relative to the estimates obtained by fitting the given model on the full data set (Full) for all scenario and sample size combinations explored.Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. Simulation scenarios incorporated trap specific behavioral effects (t2, t6, t7), individual heterogeneity in capture probability (t3, t6, t7), redundancy in sample deposition (t4, t5, t6, t7), and/or uneven distribution of activity centers (t7 and t8). Scenarios t1, t2, t3, and t8 did not include redundancy in sample deposition and did not exceed 550 samples deposited over 6 sampling periods in any simulation. All simulations included 30 individuals over 6 trapping periods. Data were subsampled using either Simple Random Sampling (SRS) or using an approach that gave preference to unique site-sessions, Site-Session Preferred (SPR).  


Figure S4. 0 versus subsampling type, scenario and model. Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. These simulation scenarios incorporated individual heterogeneity in capture probability (t3, t6, t7), redundancy in sample deposition (t4, t5, t6, t7), and/or uneven distribution of activity centers (t7 and t8; Figure 4) with 30 individuals over 6 trapping periods. Data were subsampled using either Simple Random Sampling (SRS) or using an approach that gave preference to unique site-sessions, Site-Session Preferred (SPR). Scenarios t1, t3 and t8 did not include redundancy in sample deposition and did not exceed 550 samples deposited over 6 sampling periods in any simulation.

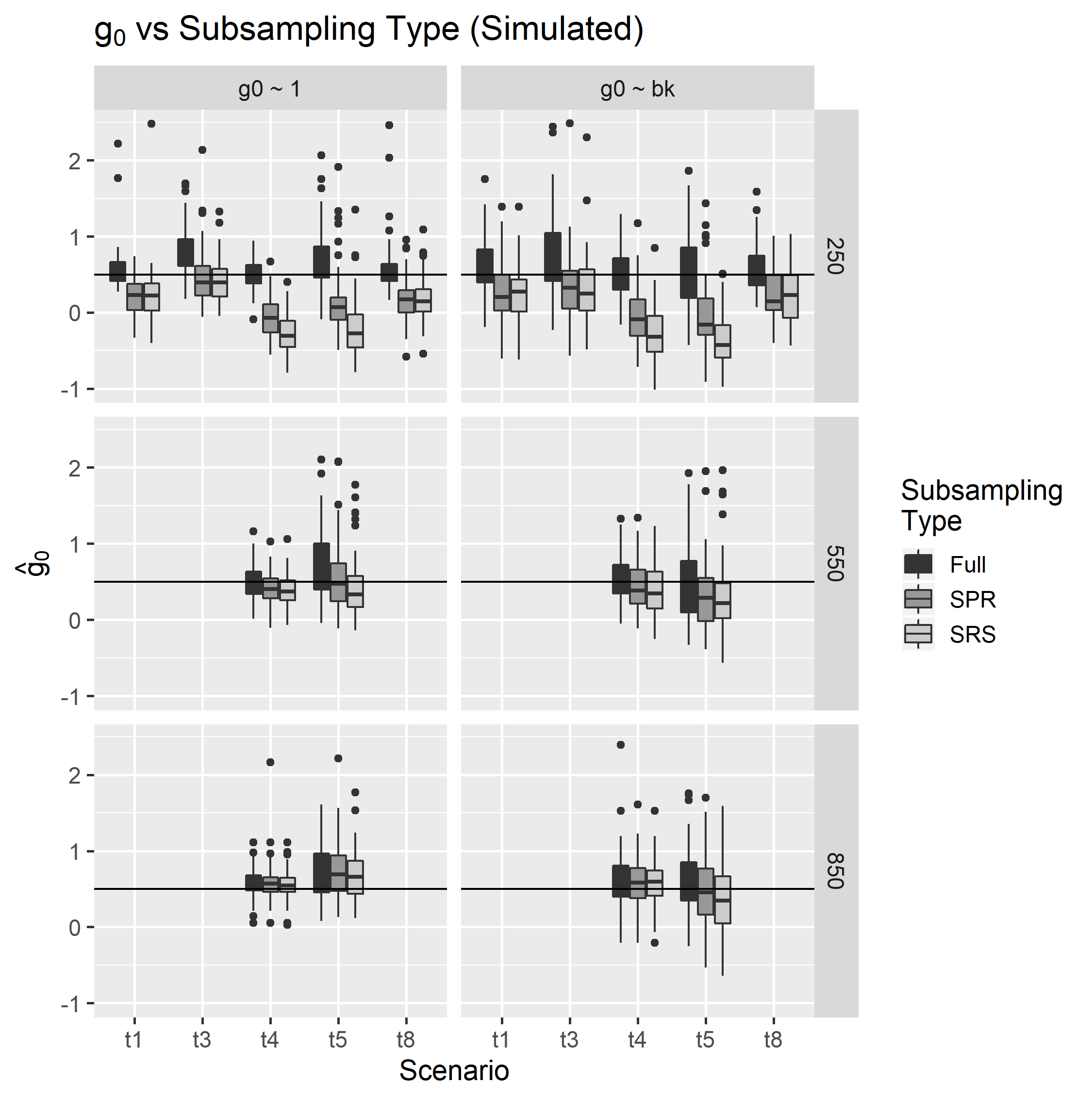
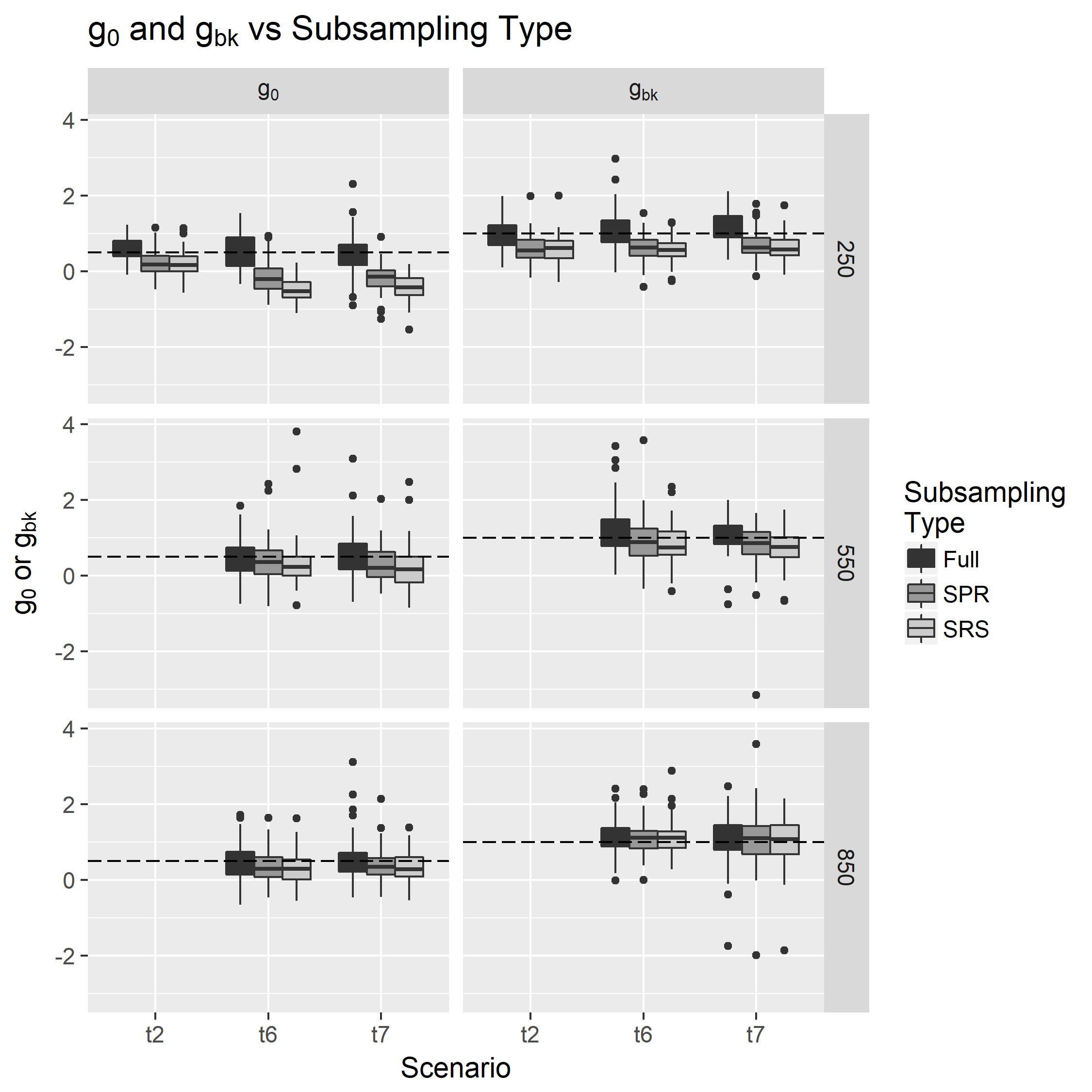


Figure S5. Initial capture probabilities at 0 distance between the activity center and trap and when δi  = 0, 0, and trap specific recapture probabilities, bk,versus subsampling type, and scenario using model g0 ~ bk.Boxes extend to the first and third quartiles of values, and whiskers extend to 1.5 times the interquartile range in either direction from the mean value. Horizontal lines represent simulation input values. Scenarios t2, t6 and t7 included a positive trap-specific behavioral effect (increased likelihood of capture at a specific trap after initial capture at that trap). All simulations included 30 individuals over 6 trapping periods. Data were subsampled using either Simple Random Sampling (SRS) or using an approach that gave preference to unique site-sessions, Site-Session Preferred (SPR). Scenario t2 did not include redundancy in sample deposition and did not exceed 550 samples deposited over 6 sampling periods in any simulation.



1. *Current affiliation: Conservation Metrics Inc, 145 McAllister Way, Santa Cruz, CA 95060 USA* [↑](#footnote-ref-1)
2. *Email*: jfieberg@umn.edu [↑](#footnote-ref-2)