Removing GPS collar bias in habitat selection studies

JACQUELINE L. FRAIR*, SCOTT E. NIELSEN*, EVELYN H. MERRILL*,
SUBHASH R. LELE†, MARK S. BOYCE*, ROBIN H. M. MUNRO‡,
GORDON B. STENHOUSE‡ and HAWTHORNE L. BEYER*

*Department of Biological Sciences and †Department of Statistics, University of Alberta, Edmonton, Alberta,
Canada T6G 2E9; and ‡Foothills Model Forest Grizzly Bear Project, Box 6330, Hinton, Alberta, Canada*

Summary

1. Compared to traditional radio-collars, global positioning system (GPS) collars provide finer spatial resolution and collect locations across a broader range of spatial and temporal conditions. However, data from GPS collars are biased because vegetation and terrain interfere with the satellite signals necessary to acquire a location. Analyses of habitat selection generally proceed without correcting for this known sampling bias. We documented the effects of bias in resource selection functions (RSF) and compared the effectiveness of two bias-correction techniques.

2. The effects of environmental conditions on the probability of a GPS collar collecting a location were modelled for three brands of collar using data collected in 24-h trials at 194 test locations. The best-supported model was used to create GPS-biased data from unbiased animal locations. These data were used to assess the effects of bias given data losses in the range of 10–40% at both 1- and 6-h sampling intensities. We compared the sign, value and significance of coefficients derived using biased and unbiased data.

3. With 6-h locations we observed type II error rates of 30–40% given as little as a 10% data loss. Biased data also produced coefficients that were significantly more negative than unbiased estimates. Increasing the sampling intensity from 6- to 1-h locations eliminated type II errors but increased the magnitude of coefficient bias. No type I errors or changes in sign were observed.

4. We applied sample weighting and iterative simulation given a 30% data loss. For a biased vegetation type, simulation reduced more type II errors than weighting, most probably because the original sample size was re-established. However, selection for areas near trails, which was influenced by a biased vegetation type, showed fewer type II errors after weighting existing animal locations than after simulation. Both techniques corrected 100% and ≥ 80% of the biased coefficients at the 6- and 1-h sampling intensities, respectively.

5. *Synthesis and applications.* This study demonstrates that GPS error is predictable and biases the coefficients of resource selection models dependant upon the GPS sampling intensity and the level of data loss. We provide effective alternatives for correcting bias and discuss applying corrections under different sampling designs.

Key-words: GPS bias, missing data, multiple imputation, radio-telemetry, resource selection function, sample weights

Journal of Applied Ecology (2004) **41**, 201–212

Introduction

Recent integration of global positioning systems (GPS) into devices for tracking animals has extended our

Correspondence: J. Frair, Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9 (fax +780 492 9234; e-mail jfrair@ualberta.ca).

ability to monitor movements of free-ranging species over a broad range of spatial and temporal conditions. Despite improvements in this technology two types of errors remain inherent in animal location data collected by GPS telemetry, namely spatial inaccuracy in the locations acquired and missing data in the form of failed location attempts. The first type of error is not unique to GPS telemetry and its effect on apparent habitat

selection has been well considered (White & Garrott 1986; Nams 1989). In particular, location inaccuracy can lead to misclassification of habitat use dependent upon the magnitude of location error and the degree of landscape heterogeneity. Location inaccuracy may be of less concern because the intentional degradation of satellite signals (selective availability) ceased in May 2000 and errors are reported to be ≤ 31 m 95% of the time (D'Eon *et al.* 2002), which is comparable to the resolution of most habitat maps. To counteract potential misclassification problems, one might resample locations within error polygons (Nams 1989; Samuel & Kenow 1992; Kenow *et al.* 2001) or replace point data with areas (buffers) around points (Kufeld, Bowden & Siperek 1987; Rettie & McLoughlin 1999).

The second type of error, missing data, has largely been ignored even though it may have a more profound effect on inferences of habitat selection than inaccurate locations (Johnson *et al.* 1998). Missing locations equate to a loss of information, the implications being reduced efficiency and potential bias in the parameters estimated by habitat selection models (Little & Schenker 1995). Bias is likely in GPS telemetry studies because failed location attempts do not occur randomly but systematically. Previous work has shown that canopy type (Moen *et al.* 1996; Moen, Pastor & Cohen 1997), percentage canopy cover (Rempel, Rodgers & Abraham 1995; Rumble & Lindzey 1997; D'Eon *et al.* 2002), tree density (Rumble & Lindzey 1997), tree height (Rempel & Rodgers 1997; Dussault *et al.* 1999) and tree basal area (Rempel, Rodgers & Abraham 1995; Rumble & Lindzey 1997) can affect the acquisition of GPS locations. For example, GPS collars have been shown to be 3·8 times less likely to acquire a location under a tall forest canopy (> 15 m height) than in treeless areas (Rempel & Rodgers 1997). In mountainous study areas, terrain conditions can interact with forest canopy cover to reduce location acquisition further (D'Eon *et al.* 2002). There are also predictable temporal effects due to the presence or absence of deciduous leaves (Dussault *et al.* 1999; Moen, Pastor & Cohen 1997) and a changing satellite constellation throughout the day (Moen, Pastor & Cohen 1997). A simulation experiment demonstrated that animal locations biased to approximate GPS error led to type II errors (failure to detect significant selection) and incorrect conclusions of selection vs. avoidance (Rettie & McLoughlin 1999). The magnitude of effects observed by Rettie & McLoughlin (1999) depended on the level of data loss, how often the animal used biased vegetation types, and the degree of spatial association among vegetation types.

Despite documentation of GPS bias, and strong recommendations for bias corrections (Rumble & Lindzey 1997; Johnson *et al.* 1998; Dussault *et al.* 1999), most statistical analyses of habitat selection continue to ignore the effects biased data may have on subsequent inferences. One suggested method for reducing these effects, in addition to the effects of spatial inaccuracy, is to measure the areal extent of each habitat type within buffers around point locations rather than the habitat type at each location (Kufeld, Bowden & Siperek 1987; Rettie & McLoughlin 1999). Using this approach, Rettie & McLoughlin (1999) were better able to identify selection vs. avoidance accurately because buffers captured portions of biased habitat types that the acquired set of locations did not. However, buffers added sampling error by including 'noise', habitats that may not affect animal behaviour, and thus their power to detect significant selection of certain habitats was reduced. Buffers therefore fail to solve the problems caused by biased missing data. Because missing GPS locations may be largely predictable, a more direct approach is to model the missing data mechanism and correct for bias statistically.

In this study, we modelled the effects of collar brand, forest structure, season, terrain and time of day on the probability of acquiring a GPS-collar location using field data. Using this model, we removed locations incrementally from an unbiased set of animal locations at two temporal sampling intensities (6- and 1-h locations). We identified the level of data loss at which coefficients in habitat selection models differed from unbiased estimates. Resource selection functions (RSF; Manly *et al.* 2002) were used to quantify selection patterns. Alternative methods exist for assessing selection, e.g. compositional analysis (Aebischer, Robertson & Kenward 1993), but we are most familiar with RSF techniques and focus solely on these. We chose a sampling design consistent with a third-order selection process (Johnson 1980), where used sites (animal locations) are compared with available sites (random locations) within the animal's home range, because this design is common to selection studies. We compared model coefficients produced using unbiased and biased data to determine how habitat-induced data loss affected the direction (selection vs. avoidance), magnitude (coefficient value) and strength (significance level) of selection. Finally, we evaluated the effectiveness of two bias-correction methods, sample weighting and iterative simulation, at removing bias from RSF coefficients. Sample weighting is a deterministic process in which the influence of each location in the data set is weighted by the inverse probability of having acquired that location (Little 1986; Kish 1992; Pfeffermann 1993). The alternative approach, iterative simulation, involves repeatedly simulating plausible spatial coordinates for each missing location and using multiple imputation methods to combine simulation results into a single model (Rubin 1987; Schafer 1999). Both techniques require a bias estimate for every location in the landscape, which we produced using field trials and data held in a geographical information system (GIS).

Materials and methods

GPS BIAS MODEL

We modelled the probability of acquiring a GPS location using data from GPS collars recorded during 194 trials in the eastern-central Rocky Mountains and foothills

Table 1. Landscape characteristics of sites where GPS collar trials were conducted, and the percentage of location attempts that were successful for three types of collars. Collars attempted locations every 30–60 min for ≥ 22 h per trial

Vegetation category	Number of trial sites	Percentage canopy mean ± SE	Percentage slope mean ± SE	Percentage location success	
				Range	Mean ± SE
Televilt GPS Simplex collars					
Non-forested	8	0·1 ± 0·0	5·8 ± 0·2	48·9–100·0	91·6 ± 6·4
Open conifer forest	8	36·8 ± 0·8	13·2 ± 0·8	53·2–98·9	85·5 ± 5·9
Closed conifer forest	11	89·7 ± 0·2	17·7 ± 0·7	12·8–92·6	67·6 ± 8·2
Deciduous forest (leaf-on)	3	83·0 ± 1·2	28·7 ± 1·0	70·2–100·0	84·4 ± 8·6
Mixed forest (leaf-on)	3	87·0 ± 0·1	15·6 ± 0·5	72·6–93·6	85·9 ± 6·7
ATS collars					
Non-forested	7	0·0 ± 0·0	13·4 ± 1·0	97·9–100·0	99·7 ± 0·3
Open conifer forest	7	29·5 ± 0·9	8·3 ± 0·3	33·3–100·0	88·0 ± 9·2
Closed conifer forest	8	90·3 ± 0·3	20·8 ± 0·7	59·4–100·0	89·5 ± 4·8
Deciduous forest (leaf-on)	1	90·0	13·7		89·6
Mixed forest (leaf-on)	1	86·0	22·2		84·4
Lotek GPS 2200 collars					
Non-forested	28	5·1 ± 0·3	12·7 ± 0·6	58·3–100·0	94·9 ± 2·6
Open conifer forest	11	47·2 ± 0·5	6·4 ± 0·4	70·8–100·0	86·7 ± 3·8
Closed conifer forest	37	84·1 ± 0·3	16·7 ± 0·5	70·2–100·0	93·5 ± 1·9
Deciduous forest (leaf-on)	11	82·7 ± 0·3	8·5 ± 0·2	50·0–100·0	87·5 ± 4·6
Deciduous forest (leaf-off)	25	83·9 ± 0·3	13·3 ± 0·4	62·5–100·0	94·3 ± 2·1
Mixed forest (leaf-on)	5	74·6 ± 1·0	3·7 ± 0·2	91·7–100·0	97·5 ± 1·7
Mixed forest (leaf-off)	22	80·2 ± 0·6	13·1 ± 0·3	60·9–100·0	96·1 ± 2·0

of Canada (52°27′N, 115°45′W). We used 10 Lotek GPS 2200 collars (2001 production; Lotek Wireless, Ontario, Canada) at 143 sites, seven Televilt GPS Simplex collars (1999 production; Televilt International, Lindesberg, Isanti, Sweden) at 33 sites, and six ATS GPS collars (2000 production; Advanced Telemetry Systems, Minnesota, USA) at 24 sites (Table 1). Logistical constraints led to uneven sample sizes among collar types. Trials were conducted from July to December 2000 and during July 2001, and consisted of placing a GPS collar approximately 1 m above ground, with the antenna directly upright, and leaving the collar to collect locations at 30- or 60-min intervals for ≥ 22 h. Trials took place across a range of conditions, from gently rolling to mountainous terrain, in open and forested areas. Forests were dominated by lodgepole pine *Pinus contorta* Dougl. ex Loud., black spruce *Picea mariana* (Mill.) B.S.P., white spruce *Picea glauca* (Moench) Voss, Engelmann spruce *Picea engelmannii* Parry ex Engelm., trembling aspen *Populus tremuloides* Michx. and balsam poplar *Populus balsamea* L.

At each trial site we recorded percentage canopy closure as the average spherical densiometer estimate across five site readings, directly over the collar and 10 m distant in the four cardinal directions. Tree height, diameter at breast height (d.b.h.) and density were recorded within a 2 × 10-m transect centred over the collar. A 100-m digital elevation model with a 30-m cell size was used to calculate terrain indices for each location using Arc/Info software (Environmental Systems Research Incorporated, Redlands, California, USA). Terrain indices included percentage slope at the test site, terrain ruggedness of the area (standard deviation in elevation within a 500-m radius) and percentage visible

sky (the amount of a hemispherical dome centred over the location that was not obstructed by terrain). Percentage visible sky was analogous to the 'available sky' index described by D'Eon *et al.* (2002). The effects of time of day have not been apparent using consecutive 4-h classes (D'Eon *et al.* 2002), most probably because various optimal and suboptimal satellite configurations can occur throughout the day. As an alternative, we pooled trials, plotted percentage location acquisition by hour, and assigned each location attempt to one of three time classes: (i) location acquisition rates > 90%, early morning (03:00–0:600), early afternoon (12:00–13:00) and evening (18:00–20:00); (ii) acquisition rates from 87% to 90%, late morning (07:00–11:00) and night (21:00–02:00); and (iii) acquisition rates ≤ 86%, late afternoon (14:00–17:00). We excluded two trials (one each for Lotek and Televilt collars) because the collars acquired < 1% of the attempted locations and we would not apply corrective measures to such obvious incidences of collar malfunction.

We used logistic regression to model the probability of a location attempt being successful (1) or unsuccessful (0) as:

$$P_{ACQ} = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)} \quad \text{eqn 1}$$

where P_{ACQ} is the probability of successfully acquiring a GPS location, β_0 is the regression constant and $\beta_1 \ldots \beta_n$ are coefficients estimated for variables $x_1 \ldots x_n$ (Hosmer & Lemeshow 2000). Because successive location attempts at a site were not independent, we used a clustering technique that recognized the unit of replication to be the trial site rather than each observation

(Pendergast *et al.* 1996; STATA Corporation 2001a). Using the techniques of Pregibon (1981), we identified several trials having high leverage but considered none to be outliers. Thus, all trials ($n = 192$) were retained for model development. We considered candidate models to be all possible combinations of non-correlated variables (Pearson $r < 0·5$ when $P < 0·05$) and appropriate interaction terms. Therefore, potential covariates included collar brand (Televilt, ATS, Lotek), season (leaf-on, leaf-off), time class, mean tree height (m), mean tree d.b.h. (cm) or percentage canopy closure, tree density (number of trees per ha) or percentage canopy closure, vegetation class (open conifer forest, closed conifer forest, deciduous forest, mixed forest, non-forested) or overstorey canopy type (open, closed, no canopy) or percentage canopy closure, and percentage slope or terrain ruggedness or percentage visible sky. For all categorical variables, we used indicator coding and selected as a reference category the class least likely to influence location acquisition.

Akaike's information criterion with a small-sample bias adjustment (AIC_c) and Akaike weights (w_i) were used to identify a set of parsimonious models that best explained our data (Burnham & Anderson 2002). From these we selected the best-supported model to calculate the probability of acquiring a GPS location across our landscape. We assessed overall model classification accuracy by using the area under the receiver operating characteristic (ROC) curve (Hanley & McNeil 1982) and model fit by the Hosmer & Lemeshow goodness-of-fit statistic ($\hat{C}$; Hosmer & Lemeshow 2000).

EFFECTS OF GPS BIAS ON RSF COEFFICIENTS

We evaluated the effects of GPS bias on habitat selection using data from a free-ranging, female wapiti *Cervus elaphus* L. inhabiting the central east slopes of the Rocky Mountains in Alberta, Canada. Actual animal locations were used to include realistic spatial and temporal autocorrelations in habitat-use patterns. We took an RSF approach to modelling habitat selection where an RSF is any statistical model that yields values proportional to the probability of resource use by an organism (Manly *et al.* 2002). The design we used for RSF estimation is commonly employed in radio-telemetry studies where characteristics of sites 'used' by animals are compared with those 'available' using logistic regression. The relative probability of animal occurrence is assumed to take the form:

$$w(x) = \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n) \qquad \text{eqn 2}$$

where $\beta_1 \ldots \beta_n$ are logistic regression coefficients estimated for environmental variables $x_1 \ldots x_n$ (Manly *et al.* 2002).

The wapiti selected for this analysis wore a Lotek collar that achieved a 96% location acquisition rate given a 1-h sampling interval over a period of 5 months, despite the animal occupying a landscape that was more

than 70% forested. Using the original data ($n = 2986$) and a resampled set of 6-h locations ($n = 497$), we estimated a RSF using three environmental variables, two of which, vegetation type and percentage slope, were variables in the best GPS bias model (P_{ACQ}). The third variable, distance to nearest trail, was included to explore the effects of both GPS bias and our corrections on variables that do not directly influence location acquisition. Vegetation type was derived from Alberta Vegetation Inventory data (Alberta Environment, Edmonton, Alberta, Canada) produced through air-photo interpretation using a 0·5-ha minimum-mapping unit and converted to a 30-m resolution grid. A grid format was required to make spatially explicit predictions of GPS error and the 30-m cell size was consistent with the resolution of Thematic Mapper satellite imagery, which is commonly used for studies on large mammals. Because the grid cell size was below the resolution of the original data there was no loss of information due to format conversion. Percentage slope was derived from the digital elevation model. Trails included 5–9-m wide recreation trails and seismic exploration transects. In this area trails occurred in each vegetation type and terrain condition proportionate to their occurrence.
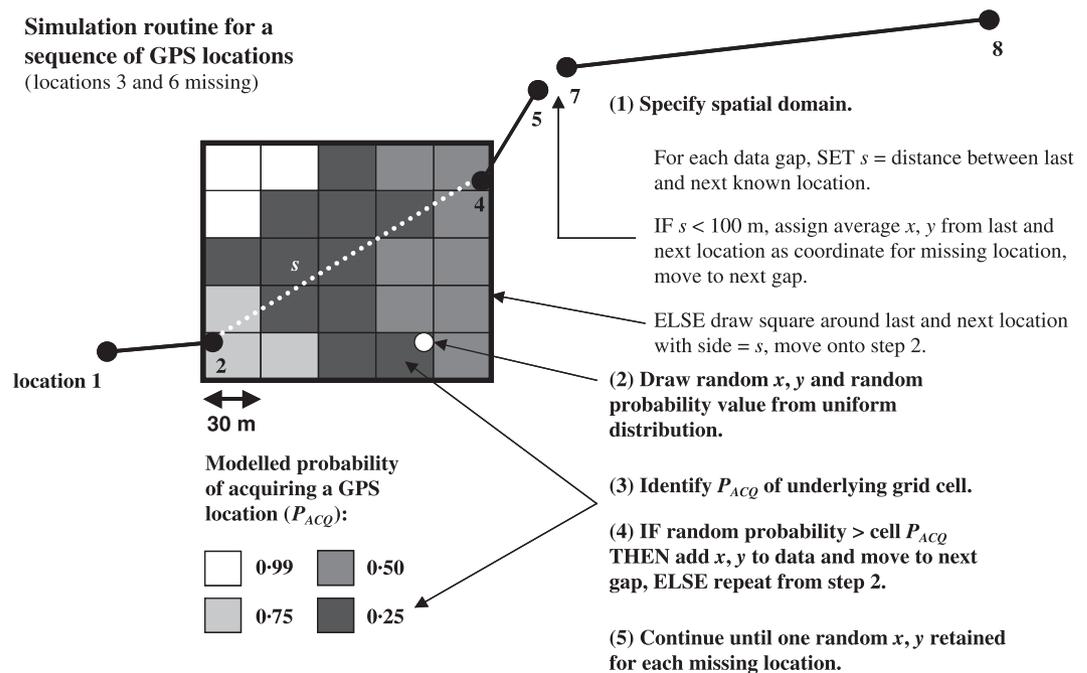
For both sampling intensities we compared RSF coefficients based on the full (unbiased) data set to subsets of these data after removing 10–40% of the locations in a biased manner. The reduction process involved randomly selecting locations for evaluation and removing a subset of selected locations according to their probability of being acquired using the P_{ACQ} model. Because the data reduction process was stochastic, we created 10 independent sets of biased data for each level of data loss. To represent resource availability we generated 2986 random locations within a minimum convex polygon (Mohr 1947) that enclosed the complete set of 1-h locations, and 497 random locations within the polygon enclosing the set of 6-h locations. The same set of available locations was used for all models produced at a given sampling intensity.

Following RSF estimation we considered the type II error rate to be the percentage of biased model coefficients that were falsely detected as non-significant when compared with the unbiased model coefficient using $\alpha = 0·05$. Likewise, type I error rates (failure to detect non-significance correctly) were determined by comparing biased to unbiased coefficients using $\alpha = 0·05$. Coverage, defined by the proportion of unbiased coefficient values that fell within the confidence intervals of the coefficients derived from biased data, was used to assess if GPS bias caused a significant change in the apparent magnitude of selection.

CORRECTIONS FOR GPS BIAS

A data loss of 30% falls at the upper end of the range of data loss reported for collars recovered from free-ranging animals (Edenius 1997; Merrill *et al.* 1998; Dussault *et al.* 1999; Biggs, Bennett & Fresquez 2001). Thus, we

Simulation routine for a sequence of GPS locations
(locations 3 and 6 missing)

(1) Specify spatial domain.

For each data gap, SET *s* = distance between last and next known location.

IF *s* < 100 m, assign average *x*, *y* from last and next location as coordinate for missing location, move to next gap.

ELSE draw square around last and next location with side = *s*, move onto step 2.

(2) Draw random *x*, *y* and random probability value from uniform distribution.

(3) Identify P_{ACQ} of underlying grid cell.

(4) IF random probability > cell P_{ACQ} THEN add *x*, *y* to data and move to next gap, ELSE repeat from step 2.

(5) Continue until one random *x*, *y* retained for each missing location.

location 1

30 m

Modelled probability of acquiring a GPS location (P_{ACQ}):

0·99	0·50
0·75	0·25

Fig. 1. Iterative simulation framework for replacing the locations missing from GPS collars.

applied two bias correction approaches given a 30% data loss to the biased 6- and 1-h location data. In the first approach, sample weighting, we applied $1/P_{ACQ}$ as a weight to each acquired location, and 1 as a weight to each available location, while estimating RSF coefficients. To calculate standard errors for coefficients we used a Huber–White sandwich estimator that is based on White's heteroscedastic-consistent estimator (White 1980; Winship & Radbill 1994; STATA Corporation 2001b).

In the second approach, iterative simulation, we 'filled in' the locations missing from each biased data set prior to estimating RSF models. The simulation process required a plausible, finite spatial domain within which each missing location was likely to have occurred (Fig. 1). For simplicity, we defined that domain to be a square centred over the last and next known animal locations. Iterative simulations required the spatial domain to contain > 2 cells; therefore, when the square domain had side length < 100 m we placed the missing location midway between the last and next known locations. We filled in each remaining missing location in a random but weighted manner using the P_{ACQ} model. Thus, we generated 30 'complete' data sets for each biased data set. We calculated a RSF for each of the 30 data sets and plotted the mean coefficient against the number of simulations conducted to discern how many iterations were needed to achieve stable estimates (Rubin 1996; Robins & Wang 2000). After selecting the necessary number of iterations, *n*, we calculated final coefficients as the average across the first *n* RSF models (Rubin 1987). The total variance associated with each coefficient was calculated as a function of the within- and between-simulation variance using multiple imputation techniques (Rubin 1987; Schafer 1999). Standard errors and sig-

nificance levels for each coefficient were calculated using a *k*-component, Student-*t* reference distribution (Barnard & Rubin 1999).

Results

COLLAR PERFORMANCE AND GPS BIAS MODEL

The mean rates of successful location attempts ranged from $67·6 \pm 8·2\%$ (SE) to $99·7 \pm 0·3\%$ across collar brands, vegetation types and terrain conditions (Table 1). Initial univariate models indicated that collar brand (Wald $\chi^2 = 11·48$, $P = 0·022$), vegetation class ($\chi^2 = 11·48$, $P = 0·022$), season ($\chi^2 = 8·54$, $P = 0·004$), tree density ($\chi^2 = 5·84$, $P = 0·016$), mean tree height ($\chi^2 = 7·92$, $P = 0·005$), percentage canopy ($\chi^2 = 3·97$, $P = 0·046$) and time class ($\chi^2 = 29·42$, $P = 0·005$) significantly affected the probability of acquiring a GPS location. The AIC_c-selected, multiple logistic regression model included collar brand, vegetation class, percentage slope and interaction terms for vegetation class × percentage slope, although there was also support for a similar model that included season (Table 2). Televilt collars had a lower probability of acquiring a GPS location than Lotek collars (the reference category), whereas ATS and Lotek collars did not differ (Table 3). Both closed conifer and deciduous forest had large and negative effects on the probability of acquiring a GPS location compared with the non-forested, reference class. The effects of open canopy conifer and mixed forest did not differ from non-forested areas. After controlling for collar brand and vegetation effects, an increasing percentage slope further reduced the likelihood of acquiring a location. However, the probability

Table 2. Comparison of the 10 highest ranked, logistic regression models for GPS bias in the eastern-central foothills of the Rocky Mountains, Alberta, Canada. The models are shown, in order of decreasing rank, with the model log-likelihood (LL), number of estimated parameters (K), Akaike's information criterion for small sample sizes (AIC_c), AIC difference (Δ_i) and AIC weight (w_i). *Interaction terms for the specified variables

Rank	Variables	LL	K	AIC_c	Δ_i	w_i
1	BRND[1], VEG[2], SLP[3], VEG*SLP	−2042·97	12	4111·66	0·00	0·55
2	BRND, VEG, SLP, VEG*SLP, SEAS[4]	−2042·02	13	4112·06	0·40	0·45
3	BRND, CAN[5], SLP, CAN*SLP	−2068·41	6	4149·27	37·61	0·00
4	BRND, VEG, SLP, SEAS	−2073·91	8	4164·60	52·94	0·00
5	BRND, VEG, SLP	−2075·79	8	4168·36	56·70	0·00
6	BRND, CAN, SLP	−2099·08	5	4208·48	96·82	0·00
7	BRND, OVER[6]	−2102·12	4	4212·45	100·79	0·00
8	STEM[7], HGHT[8], SLP	−2129·79	3	4265·70	154·04	0·00
9	STEM, HGHT	−2133·20	3	4272·53	160·87	0·00
10	BRND, VEG, SLP, VEG*SLP, HOUR	−2121·45	14	4273·25	161·59	0·00

1, Collar brand (ATS, Televilt, Lotek); 2, vegetation class (closed conifer, open conifer, deciduous, mixed forest, non-forested); 3, percentage slope; 4, season (leaf-on, leaf-off); 5, percentage canopy; 6, overstorey canopy class (closed, open, no canopy); 7, stem density; 8, tree height; 9, hour class (early morning, early afternoon and evening; late morning and night; late afternoon).

Table 3. Highest-ranked logistic regression model for predicting the probability of acquiring a GPS location (P_{ACQ}) in the in the central Rocky Mountains and foothills of Alberta, Canada (N_{obs} = 6763, Wald χ^2 = 43·70, P < 0·001, ROC area = 0·683). Standard errors were adjusted because data were clustered by trial site (n = 192)

	Adjusted			
Variable	β	SE	z	P
Vegetation type				
Open conifer forest (< 60% canopy)	−0·8515	0·6349	−1·34	0·180
Closed conifer forest (> 60% canopy)	−1·8304	0·6683	−2·74	0·006
Deciduous forest (> 60% canopy)	−1·7097	0·6379	−2·68	0·007
Mixed forest (> 40% canopy)	−0·2673	0·6906	−0·39	0·699
Reference = non-forested				
Collar brand				
ATS (2000 model)	−0·4544	0·4173	−1·09	0·276
Televilt GPS Simplex (1999 model)	−1·0969	0·2847	−3·85	< 0·001
Reference = Lotek GPS 2200 (2001 model)				
Percentage slope	−0·0316	0·0151	−2·10	0·036
Interaction terms				
Percentage slope × open conifer forest	0·0087	0·0171	0·51	0·610
Percentage slope × closed conifer forest	0·0459	0·0232	1·97	0·048
Percentage slope × deciduous forest	0·0565	0·0195	2·89	0·004
Percentage slope × mixed forest	−0·0137	0·0305	−0·45	0·654
Constant	3·8585	0·5829	6·85	< 0·001

of acquiring a location under closed conifer and deciduous forest was better on steep slopes than on flatter terrain.

Overall our bias model was significant (Wald χ^2 = 43·70, P < 0·001) and discriminated between successful and unsuccessful location attempts moderately well for Televilt (ROC area = 0·713) and ATS (ROC area = 0·664) collars. In comparison, the model poorly classified location attempts for Lotek collars (ROC area = 0·535) because these collars were highly successful at acquiring locations across the range of conditions we tested. Model predictions ranged from 0·63 to 0·98, consistent with the mean location acquisition rates observed in our trials, but did not predict the very low success observed in several trials, as reflected by the Hosmer & Lemeshow goodness-of-fit test ($\hat{C}$ = 73·76, groups = 10, n = 6693,

P < 0·001). For our simulations we solved the P_{ACQ} model for Televilt collars and recognized that our estimates for the amount of bias affecting these collars may have been conservative.

EFFECTS OF GPS BIAS ON RSF COEFFICIENTS

The unbiased selection patterns of the wapiti were the same whether 6- or 1-h locations were used, although RSF coefficients for the 1-h data were more significant ($P \leq 0·001$ excluding percentage slope) because of the larger sample size. Relative to non-forested areas, the animal avoided both closed conifer and open conifer forest and selected both deciduous and mixed forest (Table 4). The animal also selected areas close to trails while areas

Table 4. The effects of GPS-biased data loss on the detection of resource selection by a female wapiti in the Rocky Mountain foothills, Alberta, Canada. Coefficient values (β), standard errors (SE) and significance levels (P) are shown for the RSF estimated using unbiased locations collected every 6 h for 5 months ($n = 497$). Also shown are type II error rates calculated as the percentage of RSF coefficients ($n = 10$ for each level of data loss) that were falsely identified as non-significant compared with the unbiased coefficient using $\alpha = 0.05$

Variable	Unbiased model			Type II error rate after the level of data loss specified			
	β	SE	P	10%	20%	30%	40%
Vegetation type							
Closed conifer forest	−0·543	0·177	0·002	0	0	0	0
Deciduous forest	+0·534	0·263	0·042	30	20	50	40
Mixed forest	+0·639	0·243	0·008	0	0	20	70
Open conifer forest	−0·907	0·239	< 0·001	0	0	0	0
Reference = non-forested							
Distance to nearest trail (km)	−1·138	0·523	0·029	40	60	70	100
Percentage slope	+0·003	0·015	0·862	0	0	0	0

with varying percentage slope were used in proportion to their availability. Although percentage slope was not a significant variable in our unbiased RSF, we retained it to observe whether type I errors occurred due to GPS bias or our bias corrections.

No type I errors or changes in coefficient sign were observed regardless of the GPS sampling intensity or level of data loss. Likewise, for the 1-h sampling intensity, no type II errors were observed regardless of the level of data loss. For the 6-h sampling intensity, random data loss caused type II errors in the mixed forest variable once data losses reached 30%. However, type II errors due to GPS bias were prevalent in the deciduous forest and distance to nearest trail variables given as little as a 10% data loss (Table 4). A marginally significant inter-action term between closed conifer forest, a biased veg-etation type, and distance to nearest trail ($\beta = -2·37$, SE = 1·37, $P = 0·085$) indicated that areas close to trails were used more often under dense conifer canopy (all other interaction terms $P \geq 0·218$) and, thus, GPS bias indirectly affected the apparent selection of other covariates.

Biased data loss increased the magnitude of avoidance of closed conifer forest, which was significantly avoided in the unbiased model. At the 6-h sampling interval, closed conifer coefficients derived from biased data became significantly different from the unbiased coef-ficient given data losses of $\geq 30\%$ (Fig. 2). Increasing the sampling intensity from 6- to 1-h locations increased the effect of bias on the closed conifer forest variable such that 100% of the coefficients derived from biased data differed from the unbiased coefficient given a data loss of $\geq 20\%$.

EFFECTIVENESS OF BIAS CORRECTIONS

Mean coefficients for the closed conifer forest variable stabilized after 15 simulations for 6-h locations and after 25 for the 1-h locations (Fig. 3). Simulation results for the 6-h data yielded a 10% and 40% reduction of

Table 5. The effects of sample weighting and iterative simulation on detecting resource selection given a 30% biased data loss and a 6-h sampling interval. Type II error rates were calculated as the percentage of RSF coefficients ($n = 10$) that were falsely identified as non-significant when compared with the unbiased coefficient using $\alpha = 0.05$ ($\alpha = 0.10$ shown in parentheses)

Variable	Biased	Weighted	Simulated
Vegetation type			
Closed conifer forest	0 (0)	0 (0)	0 (0)
Deciduous forest	50 (10)	50 (10)	10 (0)
Mixed forest	20 (10)	20 (10)	20 (0)
Open conifer forest	0 (0)	0 (0)	0 (0)
Reference = non-forested			
Distance to nearest trail (km)	70 (40)	40 (10)	60 (20)
Percentage slope	0 (0)	0 (0)	0 (0)

type II errors in the distance to trail and deciduous forest coefficients, respectively (Table 5), and 100% coverage of the unbiased coefficients for closed conifer forest (Fig. 4a). Sample weighting reduced type II errors by 30% and 0% for the distance to trail and deciduous forest coefficients, respectively (Table 5), and also achieved 100% coverage of the unbiased closed conifer forest coefficient (Fig. 4a). Either technique combined with $\alpha = 0.10$ rather than 0·05 nearly eliminated type II errors in all variables without causing type I errors (Table 5). For the 1-h locations, simulation achieved 100% coverage and sample weighting 80% coverage of the unbiased closed conifer forest coefficient, even though coefficients were consistently underestimated (Fig. 4b). All other variables retained 100% coverage at both sampling intensities when biased and following bias corrections.

Discussion

The results from our collar tests generally agreed with previous studies in that acquisition of GPS locations was lowest under dense forest canopies, taller trees and
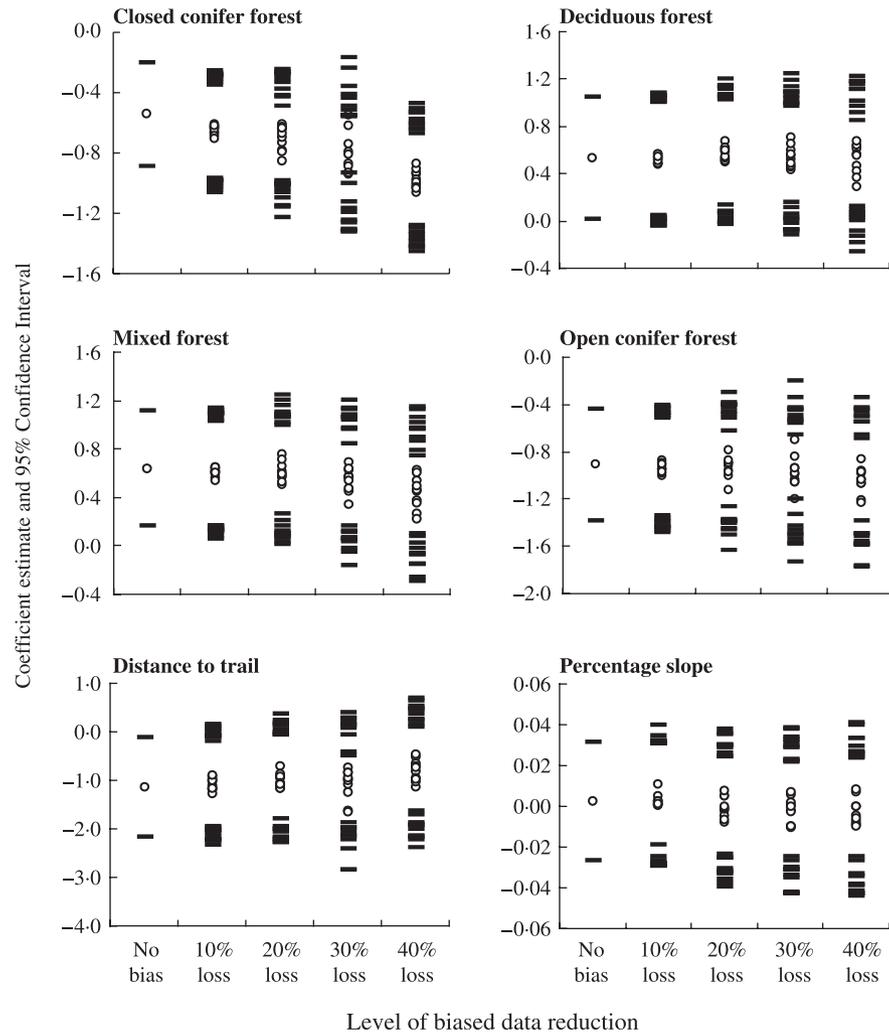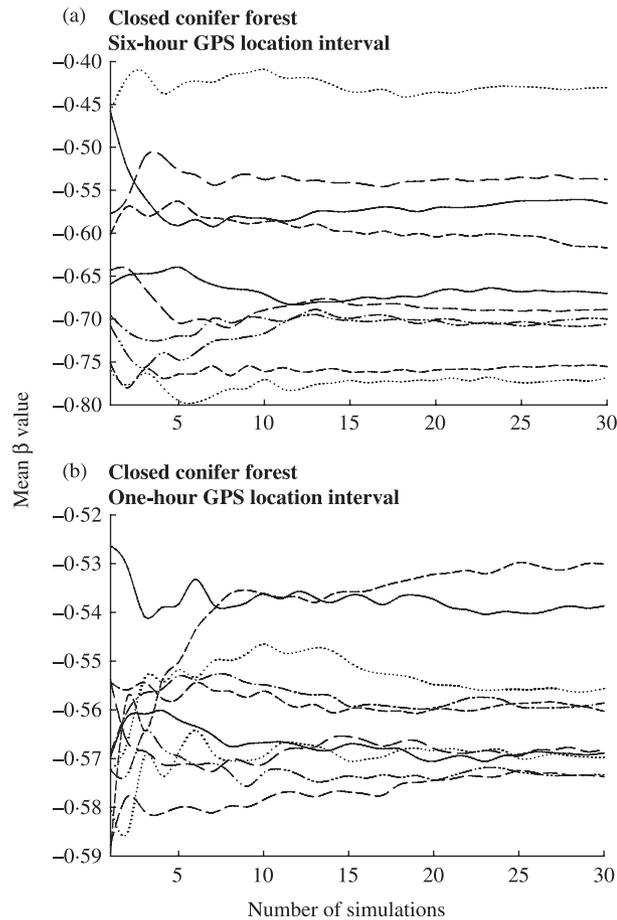
Fig. 2. The effects of biased data loss on resource selection function coefficients given a 6-h GPS location interval. Coefficient values (open circles) and 95% confidence intervals (CI; solid lines) are shown for the unbiased data model and each model produced using biased data (*n* = 10 models for each level of data reduction).

during the summer months (Moen, Pastor & Cohen 1997; Rempel & Rodgers 1997; Dussault *et al.* 1999; D'Eon *et al.* 2002). Unlike D'Eon *et al.* (2002), we found significant differences by time of day. Nevertheless, time was not a variable in our highest-ranked models and its effect on habitat selection therefore was not evaluated by our tests. We did not detect an effect of open canopy forests (< 60% canopy closure) or mixed deciduous–coniferous forest cover on location acquisition, possibly because the latter type tended to have a layered canopy with an 'open' overstorey. Terrain variables were not significant by themselves, possibly due in part to the coarse resolution of our digital elevation model. However, interactions between closed canopy forest types and percentage slope suggested that the reduction in canopy interference down-slope outweighed the potentially increased blockage of satellites up-slope due to terrain. Uncertainty among our highest-ranked models indicated that season also had important effects on GPS bias. For simplicity we did not include an effect of season in our tests but we have observed acquisition rates to vary by

season for collars recovered from free-ranging wapiti, and therefore a model including season may be necessary to compensate appropriately for GPS bias in field studies. Finally, differences in acquisition rates between collar brands may reflect, in large part, different years in which the collars were manufactured, i.e. Televilt collars were produced in 1999 and Lotek collars produced in 2001, because other researchers have reported that collar performance has improved over the years (Rempel & Rodgers 1997; Dussault *et al.* 1999).

We conclude that a GPS bias model should be produced specific to the collars employed in a given study, the specific conditions and seasons under study, and preferably produced using a sampling interval consistent with that of the free-ranging collars to be corrected. Further, animal behaviour has been shown to affect collar performance (Moen *et al.* 1996; Bowman *et al.* 2000) and collars that provide information on animal activity may additionally improve our ability to model acquisition error. We caution against extrapolating our GPS bias model to areas outside the east-central Rocky

(a) **Closed conifer forest
Six-hour GPS location interval**

(b) **Closed conifer forest
One-hour GPS location interval**

Fig. 3. Changes in the mean coefficient value for the closed conifer forest variable given the number of simulations conducted. Each line represents one of 10 sets of data, given a 30% biased data reduction, for the 6-h sampling intensity (a) and 1-h sampling intensity (b).

Mountains and foothills of Alberta because poorly fit models may introduce bias or cause excessive variation in parameter estimates (Robins, Rotnitzky & Zhao 1994). We concur with D'Eon *et al.* (2002) that unexplained or random error is a large cause of the data missing from GPS collars but, nevertheless, we have demonstrated that even a small bias resulting in small losses of data can influence our assessment of resource selection by animals.

Habitat-induced bias in animal locations acquired by GPS collars can result in type II errors and biased RSF coefficients. Several factors influenced the extent of these errors. First, rarity of certain vegetation types made them susceptible to type II errors. Similar observations have been reported by White & Garrott (1986) and Rettie & McLoughlin (1999). The two rare types, deciduous and mixed forest, were similar in extent (11% and 8% of the landscape, respectively) but deciduous forest was used slightly less (16% vs. 21%). The lower apparent strength of selection for deciduous forest ($P = 0.042$) compared with mixed forest ($P = 0.008$), combined with the large and negative effect of deciduous forest cover on GPS location acquisition, was sufficient to cause type II errors in this type given relatively small data losses (10%). Secondly, interactions among variables indicated that GPS-induced bias in one variable can influence conclusions about an animal's selection

of another resource. For example, we observed that the biased loss of locations from closed conifer forest probably caused type II errors in the distance to trail variable because the wapiti more frequently used areas near trails when under a closed conifer canopy compared with other vegetation types. Thirdly, even though closed conifer and deciduous forest had similar coefficients in the GPS bias model (Table 3), we did not observe an equivalent bias in RSF coefficients for these variables because the magnitude of use of each type of forest by the wapiti differed. Our understanding of this effect, however, differs from the simulations conducted by Rettie & McLoughlin (1999). Here closed conifer forest was the most extensive vegetation type (58% of the landscape) and was used 2·3 times more than deciduous forest, thus bias related to wapiti use of conifer forest occurred at least twice as often as for use of deciduous forest. Therefore, the magnitude of the modelled bias alone may not be sufficient to anticipate the full influence of biased data loss.

How well our corrections reduced the effects of GPS bias depended on how effectively each approach 'replaced' missing locations. Simulation increased sample sizes to their original level, thereby reducing type II errors in the rare deciduous forest type and, when combined with $\alpha = 0.10$, reducing more type II errors overall than
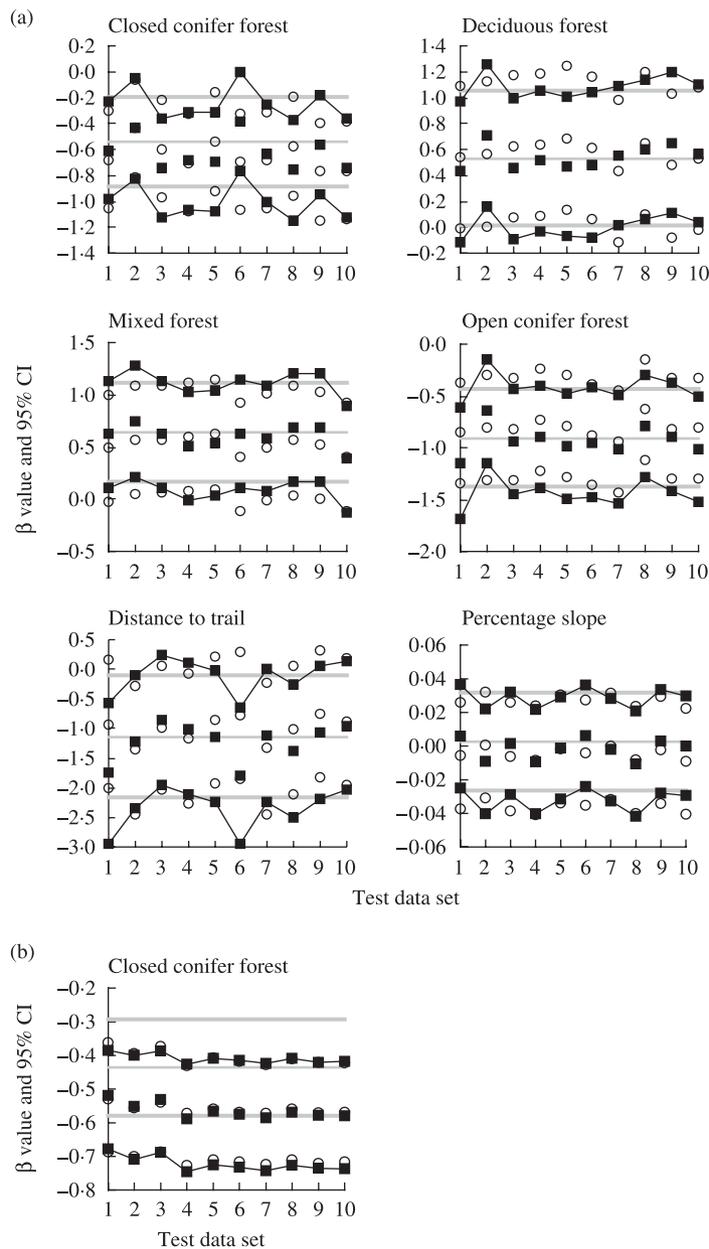
Fig. 4. The effects of sample weighting and iterative simulation on resource selection function coefficients given a 30% biased data loss. Coefficient values are shown after applying sample weights (squares) and combining simulation results (open circles) with their respective 95% confidence intervals (CI; connected squares and circles, respectively). The unbiased coefficient (thin line) and 95% confidence intervals (heavy lines) are shown for reference. For closed conifer forest, both the 6-h (a) and 1-h (b) data are shown. For the remaining variables, only the 6-h data are shown (a).

sample weighting. However, simulation placed locations on the landscape randomly with respect to trails and was thus less effective than sample weighting at removing type II errors from the distance to nearest trail variable. Sample weighting effectively 'resampled' existing animal locations, which in this case were not distributed randomly with respect to trails. Refinements to the spatial domain for imputations, e.g. limiting location replacements to within a buffer around the straight-line displacement between the last and next known locations, may better conserve the selection patterns of the animal under study and are worthy of further investigation. Further, the GPS sampling intensity

affected both the magnitude of coefficient bias and how well corrections performed. Both techniques effectively eliminated bias from closed conifer forest coefficients without introducing bias into any other variables. The extreme condition we tested of frequent sampling (1-h locations) and large data losses (30% reduction) limited our ability to correct coefficients. However, we have observed that location rates generally increase as relocation intervals shorten and thus this extreme is unlikely to be achieved in field studies.

The most suitable approach for bias correction will depend on the design for assessing resource selection. For widely roaming animals or infrequent location

schedules, sample weighting may be preferable because simulating locations within a large spatial domain may introduce an unreasonable amount of sampling error, especially in heterogeneous landscapes. Further, sample weighting may perform better than simulation when covariates are distance based (Conner, Smith & Burger 2003). Note that when sample weights are applied, a weight of one should be assigned to all influential and outlying data points to avoid unduly inflating the influence of these locations when estimating coefficients (Little & Schenker 1995). However, sample weighting may not be applicable for certain designs, such as conditional fixed-effects logistic regression, where weights cannot be applied to individual observations (Stata Corporation 2001a). For designs that temporally constrain availability (Arthur *et al.* 1996; Cooper & Millspaugh 1999; Hjermann 2000; Compton, Rhymer & McCollough 2002), iterative simulation may be more desirable as corrections are constrained to the time and area of the missed location. Further, location inaccuracy may be of concern to sample weighting as weights are applied to the GPS location rather than the true location of the animal. Our simulation routine could be adapted as part of a resampling method similar to Kenow *et al.* (2001) to account for GPS bias due to both location uncertainty and failed location attempts. Using multiple imputation techniques to combine simulation results would also be appropriate when correcting for inaccurate locations. Note that simulations should not be conducted on long sequences of missing data that occur due to random malfunction rather than GPS bias. For example, we rarely observed gaps between successful locations of greater than 8 h for Lotek collars and, thus, we used 8 h as a cut-off for corrections. Finally, both techniques support the use of point data, which overcome the limitations imposed by the use of buffers (Rettie & McLoughlin 1999). However, we have not tested the effects of bias or our corrections under any sampling design other than using logistic regression to detect a third-order selection process. We encourage exploration of bias and corrections when using any other sampling design.

Despite the increased sample sizes and increased spatial accuracy of animal locations obtained by GPS collars, inherent biases in this technology remain an evolving challenge for their users. Large-scale studies across heterogeneous landscapes may suffer unequal sample sizes among individuals due to the local effects of GPS bias. Rarification of data to investigate resource selection for specific behaviours, e.g. small- vs. large-scale movements (Johnson *et al.* 2002), or for certain time periods, e.g. day vs. night, will restrict sample sizes potentially to within the range for which we observed pervasive type II errors and coefficient bias. Further, researchers will adapt their questions to take advantage of improving technologies and, thus, sampling intervals will become increasingly shorter to the extent allowed by battery capacity. In so doing, coefficient bias may become more problematic rather than less so over time.

The bias correction techniques we present can be used to overcome many of these issues; however, large sample tests across a broad range of conditions may be necessary to understand the stability of the patterns we observed.

References

Aebischer, N.J., Robertson, P.A. & Kenward, R.E. (1993) Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, **74**, 1313–1325.

Arthur, S.M., Manly, B.F.J., McDonald, L.L. & Garner, G.W. (1996) Assessing habitat selection when availability changes. *Ecology*, **77**, 215–227.

Barnard, J. & Rubin, D.B. (1999) Small-sample degrees of freedom with multiple imputation. *Biometrika*, **86**, 948–955.

Biggs, J.R., Bennett, K.D. & Fresquez, P.R. (2001) Relationship between home range characteristics and the probability of obtaining successful global positioning system (GPS) collar positions for elk in New Mexico. *Western North American Naturalist*, **61**, 213–222.

Bowman, J.L., Kochanny, C.O., Demarais, S. & Leopold, B.D. (2000) Evaluation of a GPS collar for white-tailed deer. *Wildlife Society Bulletin*, **28**, 141–145.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York, NY.

Compton, B.W., Rhymer, J.M. & McCollough, M. (2002) Habitat selection by wood turtles (*Clemmys insculpta*): an application of paired logistic regression. *Ecology*, **83**, 833–843.

Conner, L.M., Smith, M.D. & Burger, L.W. (2003) A comparison of distance-based and classification-based analyses of habitat use. *Ecology*, **84**, 526–531.

Cooper, A.B. & Millspaugh, J.J. (1999) The application of discrete choice models to wildlife resource selection studies. *Ecology*, **80**, 566–575.

D'Eon, R.G., Serrouya, R., Smith, G. & Kochanny, C.O. (2002) GPS radiotelemetry error and bias in mountainous terrain. *Wildlife Society Bulletin*, **30**, 430–439.

Dussault, C., Courtois, R., Ouellet, J.-P. & Huot, J. (1999) Evaluation of GPS telemetry collar performance for habitat studies in the boreal forest. *Wildlife Society Bulletin*, **27**, 965–972.

Edenius, L. (1997) Field test of a GPS location system for moose *Alces alces* under Scandinavian boreal conditions. *Wildlife Biology*, **3**, 39–43.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Hjermann, D.Ø. (2000) Analyzing habitat selection in animals without well-defined home ranges. *Ecology*, **81**, 1462–1468.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression*, 2nd edn. John Wiley & Sons, New York, NY.

Johnson, B.K., Ager, A.A., Findholt, S.L., Wisdom, M.J., Marx, D.B., Kern, J.W. & Bryant, L.D. (1998) Mitigating spatial differences in observation rate of automated telemetry systems. *Journal of Wildlife Management*, **62**, 958–967.

Johnson, C.J., Parker, K.L., Heard, D.C. & Gillingham, M.P. (2002) Movement parameters of ungulates and scale-specific responses to the environment. *Journal of Animal Ecology*, **71**, 225–235.

Johnson, D.H. (1980) The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, **61**, 65–71.

Kenow, K.P., Wright, R.G., Samuel, M.D. & Rasmussen, P.W. (2001) Integrating SAS and GIS software to improve habitat-use estimates from radiotelemetry data. *Wildlife Society Bulletin*, **29**, 1006–1009.

Kish, L. (1992) Weighting for unequal P_i. *Journal of Official Statistics*, **8**, 183–200.

Kufeld, R.C., Bowden, D.C. & Siperek, J.M. (1987) Evaluation of a telemetry system for measuring habitat usage in mountainous terrain. *Northwest Science*, **61**, 249–256.

Little, R.J.A. (1986) Survey nonresponse adjustments for estimates of means. *International Statistical Review*, **54**, 139–157.

Little, R.J.A. & Schenker, N. (1995) Missing data. *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (eds G. Arminger, C.C. Clogg & M.E. Sobel), pp. 39–73. Plenum Press, New York, NY.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2002) *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*, 2nd edn. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Merrill, S.B., Adams, L.G., Nelson, M.E. & Mech, L.D. (1998) Testing releasable GPS radiocollars on wolves and white-tailed deer. *Wildlife Society Bulletin*, **26**, 830–835.

Moen, R., Pastor, J. & Cohen, J. (1997) Accuracy of GPS telemetry collar locations with differential correction. *Journal of Wildlife Management*, **61**, 530–539.

Moen, R., Pastor, J., Cohen, Y. & Schwartz, C.C. (1996) Effects of moose movement and habitat use on GPS collar performance. *Journal of Wildlife Management*, **60**, 659–668.

Mohr, C.O. (1947) Table of equivalent animal populations of North American small mammals. *American Midland Naturalist*, **37**, 233–249.

Nams, V.O. (1989) Effects of radiotelemetry error on sample size and bias when testing for habitat selection. *Canadian Journal of Zoology*, **67**, 1631–1636.

Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. & Fisher, M.R. (1996) A survey of methods for analyzing clustered binary response data. *International Statistics Review*, **64**, 89–118.

Pfeffermann, D. (1993) The role of sampling weights when modeling survey data. *International Statistical Review*, **61**, 317–337.

Pregibon, D. (1981) Logistic-regression diagnostics. *Annals of Statistics*, **9**, 705–724.

Rempel, R.S. & Rodgers, A.R. (1997) Effects of differential correction on accuracy of a GPS animal location system. *Journal of Wildlife Management*, **61**, 525–530.

Rempel, R.S., Rodgers, A.R. & Abraham, K.F. (1995) Performance of a GPS animal location system under boreal forest canopy. *Journal of Wildlife Management*, **59**, 543–551.

Rettie, W.J. & McLoughlin, P.D. (1999) Overcoming radiotelemetry bias in habitat-selection studies. *Canadian Journal of Zoology*, **77**, 1175–1184.

Robins, J.M. & Wang, N. (2000) Inference for imputation estimators. *Biometrika*, **87**, 113–124.

Robins, J.M., Rotnitzky, A. & Zhao, L.P. (1994) Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–866.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York, NY.

Rubin, D.B. (1996) Multiple imputation after 18+ years. *Journal of the American Statistician*, **91**, 473–489.

Rumble, M.A. & Lindzey, F. (1997) Effects of forest vegetation and topography on global positioning system collars for elk. *Resource Technology Institute Symposium*, **4**, 492–501.

Samuel, M.D. & Kenow, K.P. (1992) Evaluating habitat selection with radio-telemetry triangulation error. *Journal of Wildlife Management*, **56**, 725–734.

Schafer, J.L. (1999) Multiple imputation: a primer. *Statistical Methods in Medical Research*, **8**, 3–15.

STATA Corporation (2001a) *Stata 7 Reference Volume 1 (A–G)*. Stata Press, College Station, Texas.

STATA Corporation (2001b) *Stata 7 Reference Volume 2 (H–P)*. Stata Press, College Station, Texas.

White, G.C. & Garrott, R.A. (1986) Effects of biotelemetry triangulation error on detecting habitat selection. *Journal of Wildlife Management*, **50**, 509–513.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

Winship, C. & Radbill, L. (1994) Sampling weights and regression analysis. *Sociological Methods and Research*, **23**, 230–258.