

Methods for Locating African Lion Kills using Global Positioning System Movement Data

¹CRAIG J. TAMBLING, Mammal Research Institute, Department of Zoology and Entomology,
University of Pretoria, Pretoria, South Africa

ELISSA Z. CAMERON, Mammal Research Institute, Department of Zoology and Entomology,
University of Pretoria, Pretoria, South Africa

JOHAN T. DU TOIT, Department of Wildland Resources, Utah State University, Logan, Utah,
USA

WAYNE M. GETZ, Department of Environmental Science, Policy and Management, University
of California, Berkeley, California, USA

ABSTRACT Knowledge of the range, behavior, and feeding habits of large carnivores is fundamental to their successful conservation. Traditionally, the best method to obtain feeding data is through continuous observation, which is not always feasible. Reliable automated methods are needed to obtain sample sizes sufficient for statistical inference. Identification of large carnivore kill sites using Global Positioning System (GPS) data is gaining popularity. We assessed performance of generalized linear regression models (GLM) versus classification trees (CT) in a multi-predator, multi-prey African savanna ecosystem. We applied GLMs and CTs to various combinations of distance travelled data, cluster durations, and environmental factors to predict occurrence of 234 female African lion (*Panthera leo*) kill sites from 1,477 investigated

¹ E-mail: cjtambling@gmail.com

GPS clusters. Ratio of distance moved 24 hours before versus 24 hours after a cluster was the most important predictor variable in both GLM and CT analysis. In all cases, GLMs outperformed our cost-complexity-pruned CTs in their discriminative ability to separate kill from non-kill sites. Generalized linear models provided a good framework for kill site identification that incorporates a hierarchical ordering of cluster investigation and measures to assess trade-offs between classification accuracy and time constraints. Implementation of GLMs within an adaptive sampling framework can considerably increase efficiency of locating kill sites, providing a cost-effective method for increasing sample sizes of kill data.

KEY WORDS Global Positioning System (GPS), Kruger National Park, *Panthera leo*, predation, predator-prey interactions

Understanding prey-predator interactions is essential for managing African wildlife to mitigate human-wildlife conflict and to conserve prey populations (Hemson 2003, Sinclair et al. 2003, Owen-Smith et al. 2005, Grange and Duncan 2006). It is well understood that carnivore abundance is dependent upon prey availability (Smuts 1978, Gasaway et al. 1992, Mills and Funston 2003, Packer et al. 2005), which is strongly linked to climate (Ogutu and Owen-Smith 2003, 2005; Ogutu et al. 2008), and predators can regulate prey populations (Tambling and du Toit 2005, Grange and Duncan 2006, Owen-Smith and Mills 2008). However, determining relative importance of predation in regulating prey populations is challenging because predator-prey datasets commonly lack the depth and detail required for statistically rigorous analyses (Franke et al. 2006).

Hunting habits, prey selection, and range use of large African carnivores is best studied through continuous observation (Henschel and Skinner 1990, Mills 1992, Funston et al. 2001, Broomhall et al. 2003, Bissett and Bernard 2007). In African lions (*Panthera leo*, hereafter lions), continuous observation data have been used to assess factors that influence both hunting behaviour (Stander 1992a, b; Stander and Albon 1993) and success (van Orsdol 1984, Stander and Albon 1993, Funston et al. 2001), but such methods are time-consuming, labour intensive, or logistically unrealistic. Alternative techniques include opportunistic carcass location (Pienaar 1969, Schaller 1972, Mills et al. 1995, Radloff and du Toit 2004), stomach content analysis (Smuts 1979), spoor tracking (Eloff 1984), and scat analysis (Purchase 2004). All of these approaches bias the dietary results in some manner, usually by the increased detection of larger prey items (Mills 1992). Nonetheless, long term datasets have proved useful in investigating trends in lion diets in the Serengeti (Hopcraft et al. 2005), Kruger National Park (KNP; Pienaar 1969, Mills et al. 1995), and in private reserves surrounding KNP (Radloff and du Toit 2004), and methods have been suggested to correct biases (Owen-Smith and Mills 2008).

Global Positioning System (GPS) technology enables collection of high resolution spatio-temporal movement data that then can be used to locate kill sites, identify prey species, and thereby determine species-specific kill rates (Anderson and Lindzey 2003, Sand et al. 2005, Franke et al. 2006). However, this application of GPS technology is still not well developed and we know of published examples on only 2 species: pumas (*Puma concolor*: Anderson and Lindzey 2003) and wolves (*Canis lupus*: Sand et al. 2005, Franke et al. 2006, Webb et al. 2008). The primary analytic focus of these studies has been on GPS cluster aggregations in time, although an alternative could be the rate or occurrence of multiple trips to a specific location (Stotyn 2005). Technical failures of collars and time lags between event recording and data

downloading have hampered utility of GPS data in locating kills (Hemson 2002, Anderson and Lindzey 2003), although with remotely accessible data the approach is showing considerable promise in northern temperate regions (e.g., Franke et al. 2006, Webb et al. 2008).

In African savannas, where multiple predator species co-exist with multiple prey species (Hayward and Kerley 2005), complexity is added to the data and analysis because, for a given predator, residence time at a kill varies considerably with size and type of prey as well as intraguild aggression. Here we aim to develop and test models that increase the efficiency of locating lion kills from remotely accessed GPS data. These models can then be incorporated as research tools in an adaptive resource management framework (Kendall and Gould 2002).

STUDY AREA

We conducted our study in a 1,000-km² area in the central region of the KNP, South Africa, centred on Satara rest camp (31.77° E, 24.39° S). The study area was mainly open-tree savanna with a moderate to sparse shrub layer and a dense grass layer. Dominant tree species were marula (*Sclerocarya birrea*) and knobthorn (*Acacia nigrescens*) with red grass (*Themeda triandra*) and stinking grass (*Bothriocloa radicans*) dominating the grass layer (Gertenbach 1983, Venter et al. 2003). The area comprised the northern component of wildebeest (*Connochaetes taurinus*) and zebra (*Equus quagga*) migrations, resulting in high densities of these species in the wet months (Gertenbach 1983). Buffalo (*Syncerus caffer*), kudu (*Tragelaphus strepsiceros*), giraffe (*Giraffa camelopardalis*), and waterbuck (*Kobus ellipsiprymnus*) occurred in large numbers, providing a high prey density and consequently a high lion density (Gertenbach 1983, Mills and Funston 2003).

METHODS

GPS Collars and Cluster Investigation

We collared 5 female lions with GPS/global system for mobile communications (GSM) units (i.e., GPS/GSM, GPS units with mobile phone capabilities; Hawk105 units, African Wildlife Tracking ^{cc}, Pretoria, South Africa) between May 2005 and April 2007. Lions were captured using standard techniques by South African National Parks (SANParks) veterinarians (Smuts et al. 1977). Collars recorded GPS locations on 2 schedules: a) once per hour every night between 1800–0600 hours and during the day at 0900 hours, 1200 hours, and 1500 hours (16 fixes attempted) and b) once per hour over the full 24-hour period (24 fixes attempted). Collar schedules were therefore identical at night (1800–0600 hr) but differed during the day. Due to the high rate of GPS location recordings, collars had a reduced battery life and we replaced some during the study. We deployed 9 collars during the study, 4 of which attempted fixes 24 hours/day and 5 of which attempted fixes 16 hours/day. Lions in the KNP are active predominantly at night so most kills are nocturnal, with daylight hours spent resting (Mills and Biggs 1993). The combination of datasets using both collar schedules incorporated balanced sets of hourly readings at night (when lions are most active) across the entire dataset, but missed some data points during the day. The unbalanced daytime schedules should not have created any significant error due to inactivity of lions during daytime. We nevertheless performed all analyses with both schedules separately as well as with the entire pooled dataset. We downloaded data remotely via the GSM network when each collared individual entered an area of GSM coverage. We calculated distances between successive locations using the Animal Movements Extension (Hooge et al. 1999) in ArcView® 3.2 and treated missed GPS locations as stationary locations. We defined a GPS aggregation cluster (hereafter a cluster) as ≥ 2

consecutive recorded fixes with each consecutive pair of fixes <100 m apart. We used 100 m as the cut-off because over a trial period the maximum GPS error was 82 m, recorded in a riverine area. Thus, any movement >100 m can be ascribed to actual movement of lions and not GPS error. If a GPS location was not recorded and distance to the following location was >100 m, a GPS cluster was not created (even though measurement of the intervening missing points may have resulted in definition of a cluster being met).

We uploaded cluster coordinates onto a hand held GPS unit and investigated clusters on foot. We investigated an area of approximately 20-m radius around GPS points (mean GPS error prior to deployment was <20 m for all vegetation types) that encompassed all GPS points for that cluster. Therefore, we included any GPS point outside the 20-m radius of the first GPS point as the center of a new search radius so that we investigated all GPS points at the cluster. We attempted to investigate as many clusters as possible, however to maximize number of clusters investigated, we investigated clusters occurring near each other first. By conducting investigations based on proximity of many clusters, we may have biased cluster investigation to areas readily accessible by roads. We assessed this possible bias by comparing distance that clusters with or without a kill occurred from a road, as well as distance that checked and unchecked clusters occurred from a road using Wilcoxon rank sum tests. We identified predation events at clusters from presence of prey stomach contents, teeth, bones, horns, or hair and determined prey species, age, and sex when possible. The GSM coverage in the study area was not uniform and areas existed without coverage; therefore, collared females re-entered areas of coverage sporadically, resulting in an irregular pattern of cluster investigation over time (i.e., time between cluster occurrence and investigation depended on when the female moved into GSM coverage).

We separated field observations into 2 investigative periods; initially (Jun 2005 – Feb 2006) we only noted number of days between cluster occurrence and cluster investigation for kills, but subsequently (Mar 2006 – Apr 2007) we recorded number of days between cluster occurrence and cluster investigation for clusters with and without kills. During this second period, we investigated clusters between 0–671 days (\bar{x} = 54.2 days, median = 6 days) after cluster occurrence. Using Wilcoxon rank sum tests, we investigated our ability to identify species, age, and sex of the kill as time between cluster occurrence and investigation increased.

Statistical Methods

We measured 8 predictor variables for each GPS cluster. 1) Hours: length of time (hr) lions spent at a cluster from the first point of the cluster until they left the cluster for the last time. Hours was an indication of total time spent at on a carcass. 2) Return: a categorical variable describing a return visit to a cluster within 12 movement steps of leaving that cluster indicating the possibility that lions returned to carcass sites following initial movements away from the carcass site. 3) Ratio_24: ratio of distance moved during the 24 hours prior to the cluster beginning against distance moved during the 24 hours following termination of the cluster, where cluster duration included all return events. We based the variable ratio_24 on the premise that lions predominantly search for prey when hungry and rest following a kill, therefore resulting in higher ratios for successful hunts. 4) Dist_1: distance moved by lions during the first GPS interval of a cluster (i.e., between the first and second recorded GPS coordinate); a short dist_1 indicates that the female remained close to the carcass at the beginning of a cluster and did not move around the cluster site. 5) Dist_2: distance moved by lions during the second GPS interval of a cluster (i.e., between the second and third recorded GPS coordinate); a short dist_2 indicates that the female remained at the carcass, as appose to moving away from the carcass

while remaining at the cluster. 6) Drain: distance from the cluster to the nearest drainage line (classes 1-4, S.MacFadyen, KNP Scientific Services GIS department), which we calculated using ArcMap 9.0 and treated as a continuous variable, giving a measure of distance to available cover for hunting lions. 7) Road: distance from the cluster to the nearest road (S. MacFadyen, KNP Scientific Services GIS department), which we calculated as for *drain*, giving a measure of bias of investigating clusters near roads. 8) Dark: a 5-valued categorical variable that accounted for the combined effect of the sun and moon at the start of the cluster. The 5 categories were: 1 = daytime clusters, 2 = twilight clusters, 3 = night-time clusters with the moon up and full, 4 = night-time clusters with the moon up and in the first or third quarter, and 5 = night-time clusters with the moon up in the new moon phase or night-time clusters when the moon was down, as lions have been shown to have a greater kill success on dark nights

We used generalized linear models (GLMs) to investigate variables related to probability of a binary response (kill = 1, no kill = 0) occurring at a GPS cluster (Hosmer and Lemeshow 2000). We identified 2 possible time delays between cluster occurrence and cluster investigation that suggested declines in our ability to identify kills at cluster sights (see results). Therefore, we developed models based on data from clusters investigated during the first 4 weeks and first 16 weeks following a cluster occurrence. We used a forward stepwise α -to-enter approach (Quinn and Keough 2002) with a cut-off of $\alpha = 0.05$ as the criteria for entering parameters into the model. We tested parameters for collinearity and found that *hours* and *return* were correlated ($c = 0.52$), and subsequent investigation revealed that *return* was not an important variable if *hours* was already included in the model. We assessed model discrimination using the area under the curve (AUC) based on the receiver operator characteristic (ROC) curve, which is a plot of the proportion of true positives (i.e., sensitivity of discrimination) as a function of the proportion of

false positives (i.e., one minus the specificity of discrimination). With this definition the AUC score represents the percentage of time that a true positive will have a higher probability of being a kill than a true negative when both are drawn at random (Zweig and Campbell 1993, Fielding and Bell 1997). We assessed the relative individual predictors' influence on the response variable using hierarchal partitioning analysis (MacNally 2000).

Classification trees (CTs) use a hierarchal splitting criterion that separates binary responses to predictor variables based on sets of rules (Breiman et al. 1984, Ripley 2007). Threshold criteria for each variable divide responses into homogenous groups based on prior probabilities of the input data, a splitting criterion (the Gini index), and a loss matrix (Breiman et al. 1984, Ripley 2007). Although large complex trees fit the data better than small trees, large trees are not always better at predicting new data because they often over fit the data. Consequently, the construction of a best tree involves development of large trees, followed by cost-complexity pruning (Breiman et al. 1984, Ripley 2007) of branches based on honest estimates of misclassification error obtained by cross-validation (See De'ath and Fabricius 2000). As recommended by De'ath and Fabricious (2000), we ran 50 10-fold cross-validations for each tree and selected the tree that corresponded to a) the minimum relative misclassification error and b) the 1-standard error rule. Due to the unbalanced nature of our data (84% of cases were negative) we implemented a loss matrix to increase costs of a false negative error (Breiman et al. 1984). We assessed tree discrimination using AUC values for both minimum error and 1-standard error trees.

We assessed GLM and CT validation by randomly partitioning the complete dataset into independent training and testing sets and calculating their associated AUC values (Fielding and Bell 1997, Manel et al. 1999). We developed 5 model validation sets based on the k -fold

partitioning with $k = 2$ (Fielding and Bell 1997) and re-developed new GLMs and CTs (as described above) for each training set. We then tested these re-developed GLMs and CTs on the independent testing dataset. We restricted data partitioning to a 2-fold k partitioning to adhere to the suggested $M/10$ number of predictors, where M is the fewest number of cases in the binomial response (Harrell et al. 1996).

Implementation of a GLM or CT approach will be dependent on field time available for cluster investigation (i.e., more field time will enable investigation of more clusters, whereas less field time will require investigation of clusters with a higher kill probability). Therefore, we investigated the percent correct classification (PCC, all positive and negative clusters correctly identified), sensitivity (probability that a kill is correctly classified), and the specificity (probability that a non-kill is correctly classified) across a range of threshold values that enabled number of clusters investigated to vary (Fielding and Bell 1997). The threshold value converted the probability output from the model to a binary value for each cluster. Additionally, to assess how they influenced prediction success, we investigated 2 types of threshold values: 1) a default of 0.5 as the threshold and 2) the prevalence in the data as the threshold.

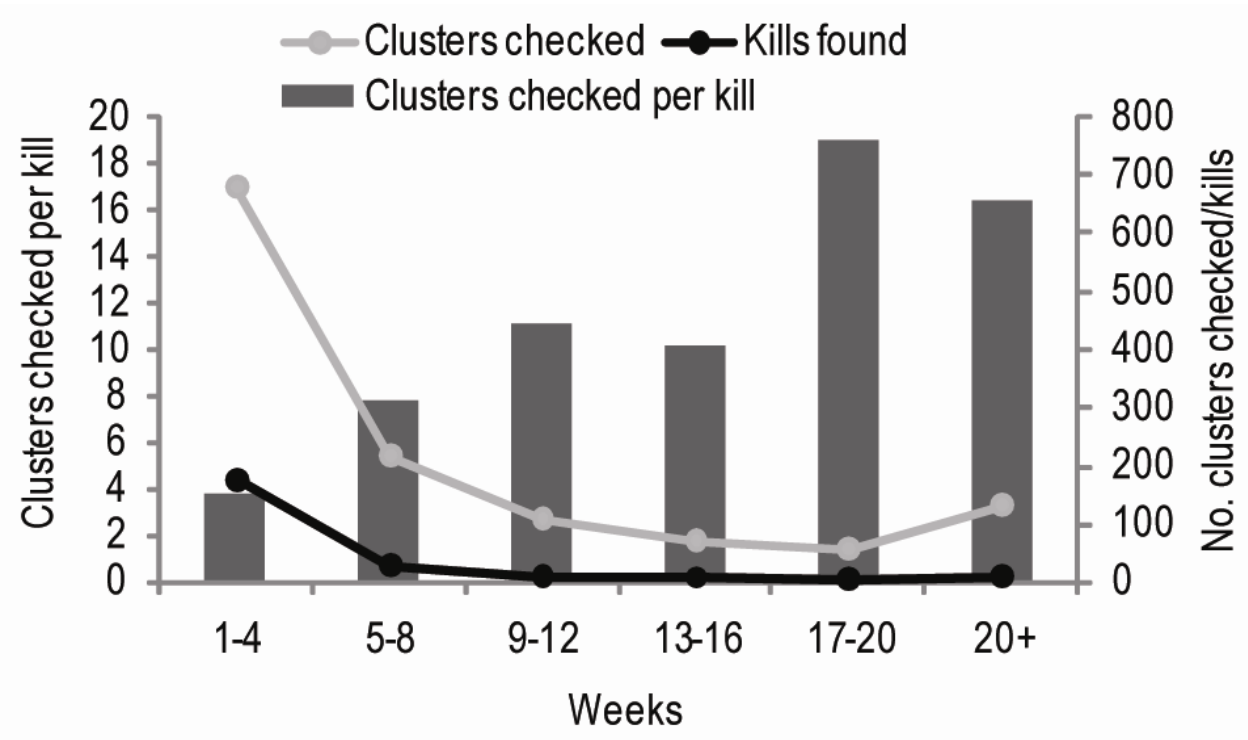
As part of an adaptive framework for kill-site prediction we developed GLMs and CTs using cumulative monthly cluster data between March 2006 and April 2007 (adaptive training datasets), and then predicted the location of kill sites for clusters investigated during the following month (adaptive testing datasets). For example, we used GLMs and CTs constructed using cluster data between March 2006 and September 2006 to predict the state of clusters investigated during October 2006. We then calculated AUC values for the testing sets. Finally, we used GLMs developed on the 5 training sets to assess effectiveness of using the resulting probability of finding a kill in the testing sets as a hierarchical means to order cluster investigation

in the field. We compared the cumulative number of kills located with each additional cluster investigated to the cumulative number of kills found by searching clusters at random and averaged it across the 5 data sets. We conducted all presence-absence and AUC analyses using the 'PresenceAbsence' package and all classification tree implementations in 'rpart' using R2.7.0 (R Development Core Team 2008).

RESULTS

Fix rate for GPS collars ranged from 65% to 88% ($\bar{x} = 77\%$). We investigated 59.5% (1,447 out of 2,433) of clusters and found 234 kills. Collared lions moved away and then returned to a cluster at 8.5% of checked clusters. Investigated clusters tended to occur closer to roads than unchecked clusters (checked clusters: 661 m, unchecked clusters: 756 m, Wilcoxon $W = 747,710$, $p < 0.05$), potentially reducing the chance of locating kills, especially because kills occurred on average farther from roads than did non kills (kills: 737 m, non-kills 647 m, Wilcoxon $W = 125,383$, $p < 0.005$). We investigated more clusters and found more kills during the first 4 weeks following cluster occurrence (673 clusters, 171 kills). Rate of clusters checked per kill remained constant during the subsequent 12 weeks. Following a 16-week interval between cluster occurrence and investigation, kills located per checked cluster declined even though sampling intensity remained the same (Fig. 1). During the 16-week subset, we found 222 kills by investigating 1,070 clusters. We found that number of days that elapsed between kills and cluster investigation did not influence our ability to confirm identification of the killed species (species confirmed: $n = 228$, species unconfirmed: $n = 6$, Wilcoxon $W = 495$, $p = 0.25$) or age classification of the prey item (confirmed: $n = 190$, unconfirmed: $n = 38$, Wilcoxon $W = 3,247.5$, $p = 0.33$). However, we did find that checking clusters sooner after cluster occurrence

Figure 1: Number of Global Positioning System (GPS) clusters investigated, number of kills found, and associated relative percentage of clusters investigated per kill during 4-week periods following occurrence of GPS clusters for female lions between May 2005 and April 2007 in Kruger National Park, South Africa.



increased our ability to classify sex of the kill, excluding juvenile kills (confirmed: $n = 50$, unconfirmed: $n = 104$, Wilcoxon $W = 2,026$, $p < 0.05$).

Preliminary model development revealed that there was no substantial improvement in model fit when the individual females were included as a random variable in a mixed-effect GLM, so we used standard GLMs for the development of predictive models. For both the 4- and 16-week datasets, 4 predictor variables were included in the final model, the most important being *Ratio_24*, explaining 51% and 48%, respectively, of the response outcome. Variables *hours* and *dark* explained 30-31% and 10-17%, respectively, of the response variable depending on dataset used, and the 3 variables collectively explained 94-95% of variation in the model

attributed to whether a kill occurred at a cluster. The final variable contributing to the explanation of kill sites was distance moved in the first GPS interval ($dist_1 = 5\text{-}6\%$). When we developed GLMs using the 2 different recording schedules the same 4 variables were selected by the α -to-enter approach and each response variable still explained a similar percentage of the response outcome.

We present CT results for the 16-week dataset only because trees developed for the 4-week dataset were similar in structure and composition. Both the 1-standard error and minimum error trees were composed of 2 splits. *Ratio_24* was the variable responsible for the main split in both CTs with kills separated from resting with a ratio >1.5 . Using the variable *Ratio_24* alone resulted in location of 64% (142/222) of kills while investigating 36% (384/1,070) of clusters, therefore locating a kill every 3 clusters checked (142 kills from 384 checked clusters). By adding the second split (min. length of a cluster >21 hr) when *Ratio_24* was <1.5 , we would have located 75% (166/222) of kills while investigating 40% (426/1,070) of clusters. Addition of a third split, which suggests investigating clusters with *Ratio_24* >0.99 when length of the cluster is <21 hours in addition to the criteria for the 2 previous splits, results in 88% (196/222) of kills located while investigating 58% (624/1,070) of clusters. As we increased number of branches our ability to locate future kills increased, but total number of clusters checked and therefore kills found declined, increasing the risk of overfitting when predicting kill sites with novel data.

The GLMs showed good discrimination with AUC values between 0.81 and 0.83 (Table 1). When we ran GLMs on the data from the 2 recording schedules AUC values remained around 0.8. Area under the curve results for the cost-complexity-pruned CTs were lower than those for the GLMs for both minimum error and 1-standard error trees, respectively (Table 1).

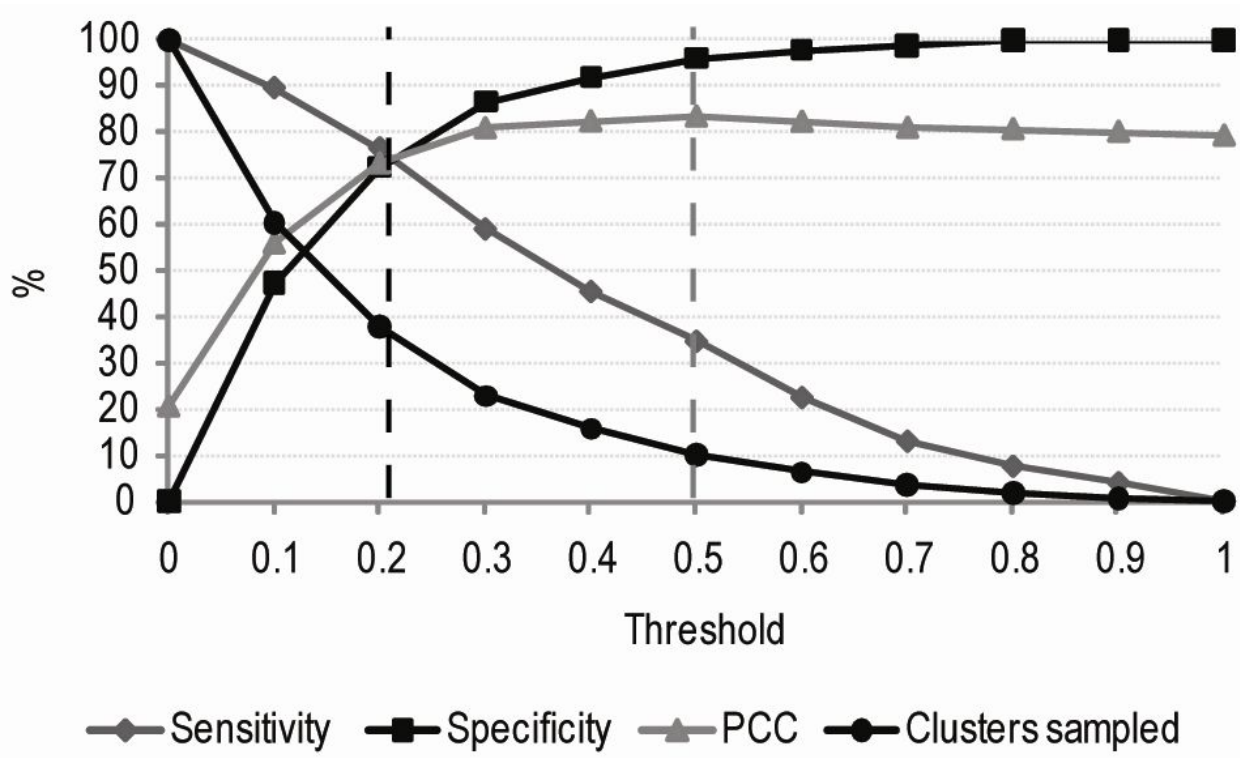
Table 1: Discriminative ability (area under the curve [AUC] of the receiver operating characteristic curve) of the generalized linear regression models (GLMs) and classification trees (CTs: including the 1-standerd error tree and min. error tree [Min]) to predict location of female lion kill sites in Kruger National Park, South Africa, between May 2005 and April 2007. Results show the discriminative ability for the full dataset against itself and the average of randomly drawn testing datasets against their associated training dataset from which we developed the models.

Data	Subset	GLM	SD	CT: Min	SD	CT: 1-SE	SD
4-week dataset	Complete	0.83		0.73		0.73	
	Testing	0.82	0.01	0.68	0.02	0.69	0.02
16-week dataset	Complete	0.81		0.72		0.72	
	Testing	0.80	0.02	0.67	0.04	0.66	0.03

Fitting the GLM of the 16-week dataset, the PCC closely followed the specificity curve in approaching the respective asymptotes above a threshold of 0.3 (Fig. 2). The PCC for a threshold value of 0.5 was 83% compared with 73% for a threshold value of 0.21 (equal to prevalence in the data). The sensitivity, however, increased from 34% to 77% when we reduced the threshold value from 0.5 to 0.21 (Fig. 2). Results for models constructed using the 4-week dataset were again similar to the 16-week dataset and we do not present them here.

Area under the curve values for testing partitions of the GLM was similar to those for the complete dataset, outperforming the cost-complexity-pruned CTs (Table 1). Similarly, from an adaptive perspective the GLM had better discriminatory ability for the following month's cluster predictions, although there was month-to-month variation (Fig. 3). Use of a hierarchal probability-based search pattern was considerably more efficient than a random search pattern for locating kill sites, locating twice as many kills after the first 200 investigated clusters (Fig. 4).

Figure 2: Sensitivity (i.e., finding a kill when a kill occurred), specificity (i.e., not finding a kill when a kill did not occur), percent correctly classified (PCC), and number of Global Positioning System (GPS) clusters we investigated for female lions between May 2005 and April 2007 in Kruger National Park, South Africa, along a range of threshold values that could be used to convert probabilities from generalized linear regression models (GLMs) to presence absence values. Two commonly used threshold points are shown by dashed lines (threshold = prevalence, 0.21: black dotted line; threshold = 0.5: grey dotted line).



DISCUSSION

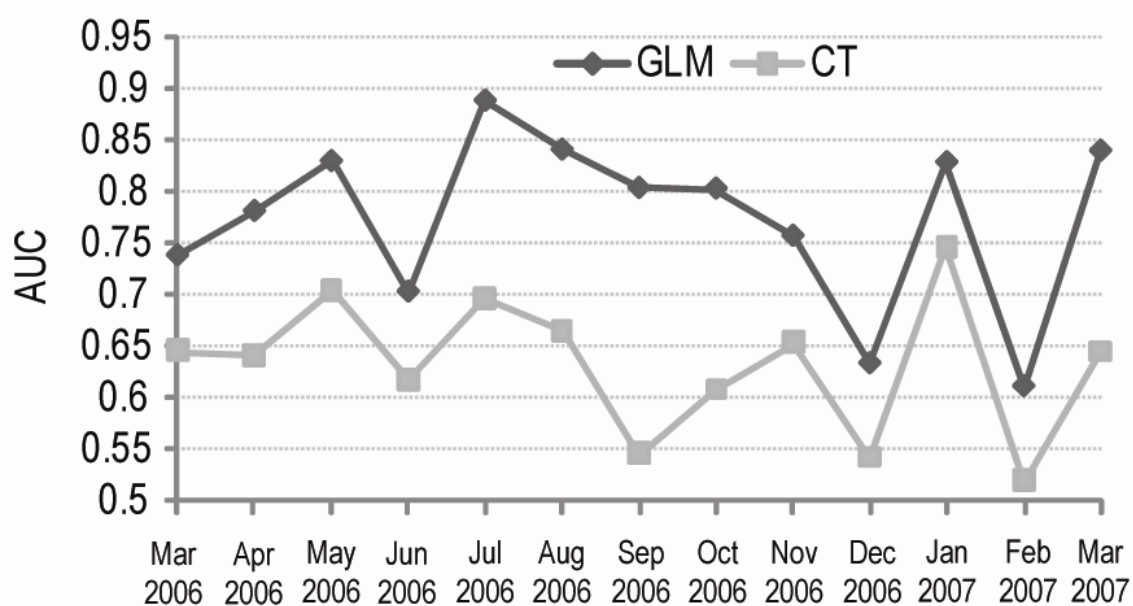
Although continual observation is the best method to investigate predator-prey relationships for large carnivores in open habitats such as African savannas (Mills 1992), it is often impossible in most other habitats (e.g., mountainous terrain, dense forests) and researchers need to use alternative methods. We found that GLMs and CTs predicted occurrence of kills at GPS clusters for female lions better than investigating clusters at random. It remains unavoidable however, that some small kills (prey items <100 kg) are missed (impala [*Aepyceros melampus*] and

warthog [*Phacochoerus africanus*] by 50%, C.J. Tambling, Mammal Research Institute, University of Pretoria, unpublished data), which is in accordance with previous studies investigating kill site detection using GPS collars (Franke et al. 2006, Webb et al. 2008).

During cluster investigation, time constraints may limit investigation to areas near roads, which could reduce number of kills located, especially in cases where predators and prey respond negatively to disturbances caused by roads (Spellerberg 1998, Kerley et al. 2002). Despite the distance to roads being non-important during model development, we did find a significant difference between distance of checked and unchecked clusters to roads. In addition, we found that clusters with kills were farther from roads than clusters without kills, therefore potentially reducing number of kills we located. We suggest that, even if time is limited, equal effort be given to investigating clusters with a high probability of being a kill at all distances from roads or access points. We expect that with increased time between cluster occurrence and investigation, false negative clusters (i.e. no kill where a kill occurred) will increase, influencing the model's predictive abilities (Vaughan and Ormerod 2005). Increasing availability of real time GPS data will assist in rapid investigation of clusters (Anderson and Lindzey 2003, Stotyn 2005). We found no difference however, in model results based on clusters investigated within 4 weeks and 16 weeks following cluster occurrence. One observer investigated >95% of all checked clusters and experience gained by this investigator enabled detection of kills that may have been missed by observers with less experience. We therefore advocate that clusters be investigated as soon as possible after they occur, as kills will be easier to locate. Because investigating clusters at random is time consuming, our results support use of a statistical model to search sites more likely to contain kills, thereby reducing field time and increasing efficiency.

Previous research suggests that hidden Markov models could be used to uncover hidden states (kill sites, bedding sites, and transit modes) from basic movement parameters (step length and turning angle) in telemetry data (Franke et al. 2006). In contrast, our analysis based on ratio of distance moved 24 hours before and 24 hours after a cluster of points proved to be the largest contributor to reliable discrimination between kill and non-kill sites. Use of this ratio as the only predictor variable however, may result in non-detection of opportunistic kills or kills made by lions employing an ambush hunting strategy (see Hopcraft et al. 2005).

Figure 3: Discriminative (area under the curve [AUC] of the receiver operating characteristic [ROC] curve) ability of the generalized linear regression model (GLM) and minimum cost-complexity classification trees (CT) to predict the following months kill sites for female lions from March 2006 until April 2007 in Kruger National Park, South Africa.



Logistic regression models have been used to predict presence or absence of kills in North American carnivores, with minimum length of time at a cluster an important predictor

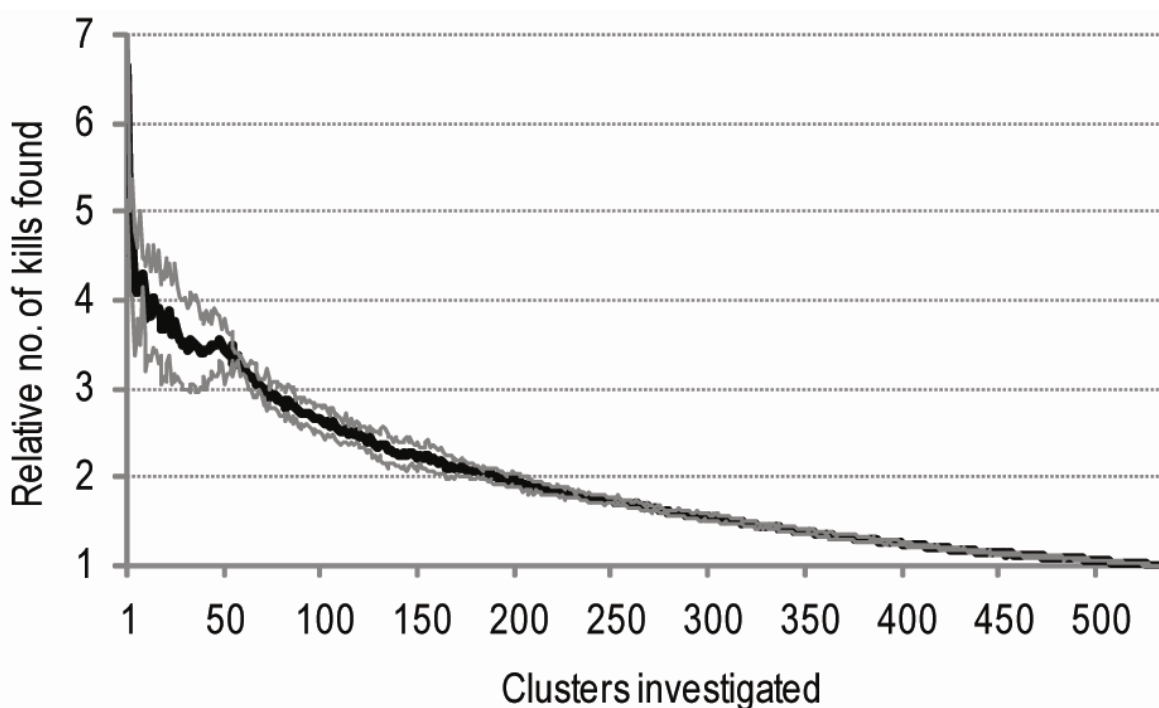
variable (Anderson and Lindzey 2003, Sand et al. 2005, Stotyn 2005, Webb et al. 2008). Our analysis revealed that, although important, length of time at a cluster alone was not enough to predict presence or absence of a kill. Handling time will undoubtedly vary between study sites (Sand et al. 2005, Webb et al. 2008) and this should be noted when developing models for new areas. Time spent at a kill site will be dependent on size of the prey item, as well as size of the group monitored. Previous investigations also suggest that occurrence of a return event to a cluster indicates presence of a kill. A low frequency of returns (8%) led to the variable *return* playing a minor role in our study, although its collinearity with overall length of time lions spent at a carcass was important, especially for large kills like adult giraffe.

Lions in KNP hunt and kill predominantly at night and hunting success tends to increase in absence of moonlight (van Orsdol 1984, Mills and Biggs 1993, Stander and Albon 1993, Funston et al. 2001). The significant influence of the darkest period in our regression models, as well as its individual importance in the hierarchal partitioning analysis, indicates its usefulness in kill site determination. However, we anticipate that incorporation of cloud cover, if known, on moonlight nights could lead to some improvement in our model.

The low importance of the GPS interval movement distances (*dist_1* and *dist_2*) might be due to the definition of the variables. Our use of the distance between the first and second, and the second and third, recorded fix of a cluster may stretch these distances on some occasions as the second and third recorded fix are not always one hour after the previous recorded GPS fix as a result of missed fixes. A stricter definition of a 1-hour movement distance at the start of the cluster, with clusters that have a missed GPS fix at the end of the first hour deleted from the dataset, may have a stronger effect on the model but will also reduce the sample size of usable

clusters. How missing data is handled needs to be carefully considered when deciding on variable use.

Figure 4: Cumulative number of kills located if we searched Global Positioning System (GPS) clusters based on probability output of the fitted GLM for all k -partitioned test datasets (grey lines represent the standard deviation for each cluster checked) relative to cumulative number of kills located if we used a random search pattern to investigate clusters for female lions in Kruger National Park, South Africa, between May 2005 and April 2007.



We used ROC curves to assess model discrimination because they are independent of threshold determination and data prevalence and provide a way of comparing among diagnostic systems (Hanley and McNeil 1982, Swets 1988, Zweig and Campbell 1993, Fielding and Bell 1997, Manel et al. 2001). The AUC values from the GLM indicated a good discriminative ability, whereas the cost-complexity-pruned CTs did not perform as well, indicating superiority of the GLM approach over CTs for predicting states associated with clusters.

The PCC of 83%, associated with a threshold of 0.5, corresponded to specificity (i.e. correct negative classification) of 96%. However, at this threshold we only investigated 10% of clusters with an expectation of finding only 34% of kills. Similarly, previous studies concerning species distribution modelling have shown an omission of known presence sites with low prevalence associated with a threshold of 0.5 (Liu et al. 2005, Jimenez-Valverde and Lobo 2007). To find more kills requires use of a lower threshold even though the resulting PCC declines. For example, when we used prevalence of the data (0.21) as a threshold, although overall PCC declined to 73%, we found an estimated 77% of all possible kills even though we only investigated 38% of clusters. Thus, in contrast to general practice (Manel et al. 2001, table 2), our study reiterates that maximizing PCC is not necessarily a good strategy. Rather, an analysis of the trade off between sensitivity and effort is needed to meet the conflicting goals of obtaining as many kills as possible within certain logistical constraints.

MANAGEMENT IMPLICATIONS

For the above approach to be properly implemented, both absence data and confirmed kills are needed for GLMs to be developed for kill site detection. To obtain absence data, we suggest investigation of low probability clusters near high probability clusters coupled with an initial period of high intensity random searching (Sand et al. 2005). From this initial data collection, models can be developed, which can improve as new data are collected from investigated clusters (Webb et al. 2008). The resulting GLM, with a sliding scale in investigative ability of researchers and a hierarchical order of cluster investigation, proved useful and therefore has merit as part of an adaptive research framework. Any combination of this approach with continual observation methods or scat collection, where possible, would refine model parameters and

increase sample sizes of kills located, resulting in increased statistical power. With advances in technology related to the study of large mammals, researchers have increasing opportunities to update traditional techniques and thus enhance efficiency of field research, thereby refining investigations of predator-prey interactions in particular.

ACKNOWLEDGMENTS

Special thanks go to A. Mabunda and K. Bulunga for their valued help and protection during fieldwork. Additional thanks to R. Dugtig who investigated GPS clusters when we were unable to make it out into the field. We thank the KNP Game Capture team, especially P. Buss and M. Kruger, for helping with immobilization and collaring of lions. We thank W. Turner and an anonymous reviewer for comments on the manuscript. This research was supported by a James S. McDonnell Foundation 21st Century Science Innovation Award, a National Institutes of Health Grant GM83863 to WMG, and a South African National Research Foundation grant to CJT.

LITERATURE CITED

- Anderson, C. R., and F. G. Lindzey. 2003. Estimating cougar predation rates from GPS location clusters. *Journal of Wildlife Management* 67:307-316.
- Bissett, C., and R. T. F. Bernard. 2007. Habitat selection and feeding ecology of the cheetah (*Acinonyx jubatus*) in thicket vegetation: is the cheetah a savanna specialist? *Journal of Zoology* 271:310-317.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Chapman & Hall, New York, New York, USA.

- Broomhall, L. S., M. G. L. Mills, and J. T. du Toit. 2003. Home range and habitat use by cheetahs (*Acinonyx jubatus*) in the Kruger National Park. *Journal of Zoology, London* 261:119-128.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.
- Eloff, F. C. 1984. Food ecology of the Kalahari lion *Panthera leo vernayi*. *Koedoe* 27 (Suppl.):249-258.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38-49.
- Franke, A., T. Caelli, G. Kuzyk, and R. J. Hudson. 2006. Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov models. *Ecological Modelling* 197:237-246.
- Funston, P. J., M. G. L. Mills, and H. C. Biggs. 2001. Factors affecting the hunting success of male and female lions in the Kruger National Park. *Journal of Zoology, London* 253:419-431.
- Gasaway, W. C., R. D. Boertje, D. V. Grangaard, D. G. Kelleyhouse, R. O. Stephenson, and D. G. Larsen. 1992. The Role of Predation in Limiting Moose at Low Densities in Alaska and Yukon and Implications for Conservation. *Wildlife Monographs* 120:3-59.
- Gertenbach, W. P. D. 1983. Landscapes of the Kruger National Park. *Koedoe* 26:9-121.
- Grange, S., and P. Duncan. 2006. Bottom-up and top-down processes in African ungulate communities: resources and predation acting on the relative abundance of zebra and grazing bovids. *Ecography* 29:899-907.
- Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36.

- Hayward, M. W., and G. I. H. Kerley. 2005. Prey preferences of the lion (*Panthera leo*). *Journal of Zoology*, London 267:309-322.
- Hemson, G. 2002. The potential for using GPS-collars on wild African lion (*Panthera leo*); some lessons from experience. *in* A. J. Loveridge, T. Lynam, and D. W. Macdonald, editors. *Lion Conservation Research*. Wildlife Conservation Research Unit, Oxford, United Kingdom.
- Hemson, G. 2003. *The Ecology and Conservation of Lions: Human-Wildlife Conflict in semi-arid environments*. Dissertation, University of Oxford, Oxford, United Kingdom.
- Henschel, J. R., and J. D. Skinner. 1990. The diet of the spotted hyaena (*Crocuta crocuta*) in Kruger National Park. *African Journal of Ecology* 28:69-82.
- Hooze, P. N., W. Eichenlaub, and E. Solomon. 1999. *The animal movement program*. Alaska Biological Science Centre, Anchorage, Alaska, USA.
- Hopcraft, J. G. C., A. R. E. Sinclair, and C. Packer. 2005. Planning for success: Serengeti lions seek prey accessibility rather than abundance. *Journal of Animal Ecology* 74:559-566.
- Hosmer, D. M., and S. Lemeshow. 2000. *Applied Logistic Regression*. John Wiley & Sons, New York, New York, USA.
- Jimenez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica* 31:361-369.
- Kendall, W. L., and W. A. Gould. 2002. An appeal to undergraduate wildlife programs: send scientists to learn statistics. *Wildlife Society Bulletin* 30:623-627.
- Kerley, L. L., J. M. Goodrich, D. G. Miquelle, E. N. Smirnov, H. B. Quigley, and M. G. Hornocker. 2002. Effects of roads and human disturbance on Amur tigers. *Conservation Biology* 16:97-108.

- Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385-393.
- MacNally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – ‘predictive’ and ‘explanatory’ models. *Biodiversity and Conservation* 9:655–671.
- Manel, S., J.-M. Dias, and S. J. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120:337-347.
- Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921-931.
- Mills, M. G. L. 1992. A comparison of methods used to study food habits of large African carnivores. *in* D. R. McCullough, and R. H. Barrett, editors. *Wildlife 2001: Populations*. Elsevier Science Publishers LTD, London, United Kingdom.
- Mills, M. G. L., and H. C. Biggs. 1993. Prey apportionment and related ecological relationships between large carnivores in Kruger National Park. *Symposium of the Zoological Society of London* 65:253-268.
- Mills, M. G. L., H. C. Biggs, and I. J. Whyte. 1995. The relationship between rainfall, lion predation and population trends in African herbivores. *Wildlife Research* 22:75-88.
- Mills, M. G. L., and P. J. Funston. 2003. Large carnivores and savanna heterogeneity. Pages 370-388 *in* J. T. du Toit, K. H. Rogers, and H. Biggs, editors. *The Kruger Experience: Ecology and Management of Savanna Heterogeneity*. Island Press, Washington DC, USA.

- Ogutu, J. O., and N. Owen-Smith. 2003. ENSO, rainfall and temperature influences on extreme population declines among African savanna ungulates. *Ecology Letters* 6:412-419.
- Ogutu, J. O., and N. Owen-Smith. 2005. Oscillations in large mammal populations: are they related to predation or rainfall? *African Journal of Ecology* 43:332-339.
- Ogutu, J. O., H.-P. Piepho, H. T. Dublin, N. Bhola, and R. S. Reid. 2008. Rainfall influences on ungulate population abundance in the Mara-Serengeti ecosystem. *Journal of Animal Ecology* 77:814-829.
- Owen-Smith, N., D. R. Mason, and J. O. Ogutu. 2005. Correlates of survival rates for 10 African ungulate populations: density, rainfall and predation. *Journal of Animal Ecology* 74:774-788.
- Owen-Smith, N., and M. G. L. Mills. 2008. Predator-prey size relationships in an African large-mammal food web. *Journal of Animal Ecology* 77:173-183.
- Packer, C., R. Hilborn, A. Mosser, B. M. Kissui, M. Borner, G. Hopcroft, J. Wilmshurst, S. Mduma, and A. R. Sinclair. 2005. Ecological change, group territoriality, and population dynamics in Serengeti lions. *Science* 307:390-393.
- Pienaar, U. d. V. 1969. Predator-prey relationships amongst the larger mammals of the Kruger National Park. *Koedoe* 12:108-176.
- Purchase, G. K. 2004. Factors affecting the ratio of lions (*Panthera leo*) to spotted hyaenas (*Crocuta crocuta*) in protected areas of Africa: Competition and/or available prey. Dissertation, University of Aberdeen, Aberdeen, Scotland.
- Quinn, G. P., and M. J. Keough. 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge, United Kingdom.

- R Development Core Team. 2008. R: A Language and Environment for Statistical Computing, 2.7.0. R Foundation for Statistical Computing. Vienna, Austria. <<http://www.R-project.org/>>. Accessed 15 Mar 2008.
- Radloff, F. G. T., and J. T. du Toit. 2004. Large predators and their prey in a southern African savanna: a predator's size determines its prey size range. *Journal of Animal Ecology* 73:410-423.
- Ripley, B. D. 2007. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, United Kingdom.
- Sand, H., B. Zimmermann, P. Wabakken, H. Andren, and H. C. Pedersen. 2005. Using GPS technology and GIS cluster analyses to estimate kill rates in wolf-ungulate ecosystems. *Wildlife Society Bulletin* 33:914-925.
- Schaller, G. B. 1972. *The Serengeti Lion: A Study of Predator-Prey Relations*. The University of Chicago Press, Chicago, Illinois, USA.
- Sinclair, A. R. E., S. Mduma, and J. S. Brashares. 2003. Patterns of predation in a diverse predator-prey system. *Nature* 425:288-290.
- Smuts, G. L. 1978. Interrelations between predators, prey, and their environment. *Bioscience* 28:316-320.
- Smuts, G. L. 1979. Diet of lions and spotted hyaenas assessed from stomach contents. *South African Journal of Wildlife Research* 9:19-25.
- Smuts, G. L., I. J. Whyte, and T. W. Dearlove. 1977. A mass capture technique for lions. *East African Wildlife Journal* 15:81-87.
- Spellerberg, F. I. 1998. Ecological effects of road and traffic: a literature review. *Global Ecology and Biogeography Letters* 7:317-333.

- Stander, P. E. 1992a. Cooperative hunting in lions: the role of the individual. *Behavioural Ecology and Sociobiology* 29:445-454.
- Stander, P. E. 1992b. Foraging dynamics of lions in a semi-arid environment. *Canadian Journal of Zoology* 70:8-21.
- Stander, P. E., and S. D. Albon. 1993. Hunting success of lions in a semi arid environment. *Symposium of the Zoological Society of London* 65:127-143.
- Stotyn, S. 2005. Determining wolf summer predation patterns using GPS cluster analysis: a preliminary report. Department of Renewable Resources, University of Alberta. Edmonton, Canada.
- Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240:1285-1293.
- Tambling, C. J., and J. T. du Toit. 2005. Modelling wildebeest population dynamics: implications of predation and harvesting in a closed system. *Journal of Applied Ecology* 42:431-441.
- van Orsdol, K. G. 1984. Foraging behaviour and hunting success of lions in Queen Elizabeth National Park, Uganda. *African Journal of Ecology* 22:79-99.
- Vaughan, I. P., and S. J. Ormerod. 2005. The continuing challenges of testing species distribution models. *Journal of Applied Ecology* 42:720-730.
- Venter, F. J., R. J. Scholes, and H. C. Eckhardt. 2003. The Abiotic Template and its Associated Vegetation Patterns. Pages 83-129 in J. T. du Toit, K. H. Rogers, and H. C. Biggs, editors. *The Kruger Experience: Ecology and Management of Savanna Heterogeneity*. Island Press, Washington DC, USA.

Webb, N. F., M. Hebblewhite, and E. H. Merrill. 2008. Statistical methods for identifying wolf kill sites using global positioning system locations. *Journal of Wildlife Management* 72:798-807.

Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39:561-577.

Associate Editor: Bowman.