Application of Distance Sampling for Pima Pineapple Cactus (*Coryphantha scheeri* var. *robustispina*) Population Enumeration and Monitoring

FINAL REPORT

Arizona Department of Agriculture
Endangered Species Act – Section 6 Grant Program
Project No. - Segment 19, 2015-2017-04

Presented to:
Deborah Atkinson
Arizona Department of Agriculture
1688 W. Adams St.
Phoenix, Arizona 85007
datkinson@azda.gov

Authors:
Aaron D. Flesch
A.D. Flesch and Associates
1116 W. Alameda Street
Tucson Arizona, 85745
flesch@ag.arizona.edu

Ian Murray and Brian Powell
Pima County Office of Sustainability and Conservation
201 N. Stone Ave
Tucson, AZ 85701

September 30, 2017

ABSTRACT

Efficient and accurate surveys methods are essential for understanding abundance and habitat relationships of plants and wildlife, especially species of conservation concern. The Pima Pineapple Cactus (*Coryphantha scheeri* var. *robustispina*) is a federally-listed endangered species in southern Arizona that is commonly surveyed for compliance with federal law and for research and monitoring. We tested a new survey method for this species based on distance sampling (DS), which involves measuring distances to focal objects from lines or points and modeling a detection function that adjusts estimates of abundance for variation in detection probability. We compared estimates of density and population size from DS with values obtained with the recommended survey protocol (Roller method) that focuses on a complete census of all individuals in a focal area. Additionally, we assessed factors that influenced detection probability and detection distance, associations between local estimates of density from DS and various environmental factors, and summarized recommendations for future applications of DS in this system. We recorded 105 live and 15 dead Pima Pineapple Cacti while DS along 36.9 km of line transects at 11 focal sites in the Brawley and Santa Cruz watersheds. Density averaged 1.47 live individuals/ha with an estimated total of 294 individuals overall based on DS, and precision was high (CV= 0.139). Based on what we presumed to be known values from census efforts, both density and abundance were fairly well estimated by DS with bias equaling -11.4% across the population. Moreover, estimates from DS were highly correlated with values from censuses on both the untransformed (r = 0.82, p = 0.002), and especially, logarithmic scales (r = 0.92, p < 0.001). Estimates of detection probability averaged 0.49 (CV = 0.076) overall, and 0.96 and 0.92 at 2 and 3 m from lines, respectively, suggesting results based on the Roller method fail to detect 4-8% of individuals on average. Detection probability declined as soils became increasingly dominated by larger rocky particles compared to those dominated by sand or silt. Detection probability also seemed to decline with increasing understory vegetation volume, grass cover, and decreasing cactus height, although additional effort is needed to confirm these patterns. Local densities increased with increasing slope and soil substrate size, and decreased with increasing understory vegetation volume ($p \leq 0.022$). There was some evidence that bias of estimates among focal sites increased as grass cover increased, suggesting more focused survey effort in areas of dense grass near lines will improve accuracy of DS. Although our results indicate DS is an effective method for surveying this species for research and monitoring applications, reducing the spacing of neighboring line transects from 50 to 40 m, surveying perpendicular to elevation contours, and other recommendations we present should improve accuracy.

INTRODUCTION

The Pima Pineapple Cactus (*Coryphantha scheeri* var. *robustispina*; hereafter “PPC”) has a relatively narrow distribution near the ecotone between desert-scrub and semi-desert grasslands in south-central Arizona and adjacent Sonora, Mexico. In response to various threats such as urban development, invasion of non-native grasses, overgrazing, and climate change, the PPC was listed as endangered in 1993 and a draft recovery plan was recently issued by the U.S. Fish and Wildlife Service (USFWS 2007, 2017). Although a number of studies provide information on the ecology and distribution of the PPC (e.g., Roller 1996a; McDonald 2005; Kidder 2015), major gaps of knowledge remain. Among these information gaps, is the need for an efficient survey method that can provide inferences on the distribution and abundance of this endangered species across both space and time, and information on environmental factors that influence these parameters.

With regard to survey methods, the current recommended survey protocol for the PPC (Roller 1996b) focuses on complete enumeration of population size in a focal area. Such an approach is appropriate where ground disturbance is proposed or in other contexts where complete population enumeration and mapping are required for compliance with the Endangered Species Act. Nonetheless, efforts to census a given population are likely too time intensive to be efficient for most research and monitoring applications, and may be based on unrealistic assumptions of perfect detection probability. These other applications include efforts to assess patterns of distribution and abundance across space and time, and habitat relationships. In such contexts, robust field methods based on sound sampling theory should be capable of providing valid inference with greater efficiency.

Distance sampling (DS) is an effective method for estimating the distribution, abundance, and habitat relationships of both terrestrial and aquatic populations of wildlife, and is applicable over broad geographic areas (Thomas et al. 2002, 2010). Although DS has mainly been applied to vertebrate populations (e.g., Rosenstock et al. 2002, Flesch et al. 2016), it can be useful for plants despite few examples of its application (Buckland et al. 2007, Kissa and Sheil 2012, Schorr 2013). Distance sampling involves measuring the distances to focal objects from lines or points, and modeling a detection function that adjusts estimates of abundance for variation in detection probability. A recent pilot effort focused on applying DS to measuring abundance of the PPC produced encouraging results (Powell 2015). Before DS can be adopted to address research and monitoring questions for the PPC, however, more information and testing are needed. Although DS is highly efficient for estimating and monitoring spatiotemporal variation in abundance, it is largely untested in arid environments and for plants (but see Anderson et al. 2001).

Here, we test a novel field-based approach for surveying the PPC based on DS. To do so, we compare estimates of population size and densities obtained during DS procedures with estimated known values of these parameters based on censuses, which were often combined with intensive repeated monitoring. We also discuss issues related to the design and implementation of DS for the PPC based on our findings, and assess factors that influence detection probability during line-transect sampling. Additionally, we assess associations between local estimates of density of PPC (e.g., within-site estimates) and environmental factors such as vegetation cover and soil substrate size to provide an example of how DS can be applied to understand plant-habitat relationships.

OBJECTIVES

Our project focused on the following three objectives:

1. A statistical comparison of estimates of PPC density and abundance based on DS and the Roller (1996b) or other similar census methods in the same areas.

2. Estimates of the influence of environmental variables on abundance and detection probability of PPC.
3. Guidance on the application of DS for PPC monitoring and research programs.

METHODS

Study Area and Design: We surveyed the PPC at sites in the Brawley (Altar Valley) and Santa Cruz watersheds in eastern Pima County located south of Tucson, Arizona (Fig. 1). We focused in areas known to support PPC and selected study sites based largely on two criteria: 1) areas with long histories of intensive PPC monitoring such that abundances and densities of PCC were largely known and had been documented within approximately 4 years of our efforts, and 2) lands owned or managed by Pima County where the presence of PPC was known but where densities were unknown and thus needed to be documented to support our efforts. In addition to these criteria, we stratified effort between the two dominant vegetation communities (Sonoran desert-scrub and semi-desert grassland) in which PPC occur, and sampled across gradients in elevation and the natural range of variation in PPC densities commonly found in Arizona.

With regard to the first criteria, we selected 5 of 6 sites (Anvil, Guy Street, Mendoza, Palo Alto, Stagecoach) in the Brawley watershed where long-term PPC monitoring efforts were established in 1997 by B. Schmalzel (2000) and 2002 by R. Routson (2003) and continued by M. Baker through September 2012 (Baker 2013). We also considered 2 sites on Sycamore Canyon Properties east of Sahuarita where PPC monitoring began in 2004 and continued until just before our surveys (Westland Resources 2004, 2017, S. Hart, pers. comm.). At each of these 7 sites, the distribution and abundance of PPC were initially documented using the Roller (1996b) method. This approach involves multiple observers spaced 2-3 m apart walking parallel transects and exhaustively searching for cacti until a focal area is completely covered, with additional effort recommended in some situations. After initial site surveys, PPC that were found were monitored across time, which involved observers searching for previously unknown cacti while walking new routes to known plants so as to maximize coverage, and adding new individuals to the sample. An untested assumption of the Roller method is that it allows for complete enumeration of population size at a focal site because detection probability is assumed to be perfect within 2-3 m of observers. Access to all sites was provided by landowners and land management agencies.

With regard to the second criteria, we selected 4 areas, including one on Canoa Ranch and 3 on Sopori Ranch, as additional study sites (Fig. 1). These sites were selected to augment sample sizes for comparing known and distance-based estimates of densities, to ensure effort spanned the natural range of variation of PPC densities in the wild, and to bolster inferences on associations between densities and environmental factors. At these sites, 2 observers (different from those that completed DS) employed the Roller method in an attempt to completely enumerate population size. Additionally, any new individuals discovered incidentally after surveys during DS were incorporated into estimates of population size at each site. Thus, assuming the accuracy of past surveys and population stability, PPC populations at sites we sampled were completely enumerated during that last 0-4 years so that estimates from DS could be compared to estimated known values.

Before DS, we uploaded plot boundaries into handheld global positioning systems (GPS) to ensure surveys overlapped past coverage. Because >80% of individuals occupied only a portion of the Anvil and Mendoza plots (4 of 5 for Anvil; 69 of 71 for Mendoza), we implemented DS on portions of these plots and adjusted Baker's estimates (2013) to improve efficiency. We used recent pilot data from DS the PPC along line transects (Powell 2015) to guide survey design. That effort found an effective strip width of transects between 8 and 13 m, and that the furthest PPC detected from lines was at ≈25 m. Thus, we placed sets of parallel line transects 50-m apart across plot boundaries in a direction approximately equaled to the longest dimension of each plot (except at Sycamore Canyon A where transects were parallel to shortest plot dimension), and began surveys from a random point on plot boundaries approximately 25 m from the edge of the boundary. We then sampled environmental conditions (see below) at 100-m intervals along each line, beginning approximately 100 m from plot boundaries (Fig. 1).

Distance-sampling surveys: Distance sampling stationary objects such as plants involves two main assumptions to ensure accurate estimation: 1) perfect detection of focal objects on the transect line, and 2) accurate measurement of distances between lines and focal objects. Additionally, a key design consideration when implementing DS is to place lines according to a randomized design. This ensures lines are positioned independently of focal objects and that objects are not uniformly distributed with respect to their distances from lines, which can bias estimates (Buckland et al. 2015).

To implement DS, two surveyors slowly walked each line with one surveyor focused on and immediately around lines, while the other surveyor walked short serpentine paths within approximately 0-6 m of lines (see Fig. 3 in Anderson et al. 2001). This arrangement ensured focused effort on and around the center line as well as effort along both sides of lines. Surveyors carefully scanned clumps of vegetation focusing near center lines to help ensure all PPC on or immediately around lines were detected. Observers also frequently scanned behind them to ensure cacti that may have been obstructed in one direction were detected from the opposite direction. Because PPC sometimes occurred in small clusters of several individuals spaced ≈10-20 m apart, after detecting a PPC observers scanned the surrounding area from lines, and noted individuals detected only while measuring cacti away from lines as incidentals, which were not included in analyses. All surveys were conducted during low winds (<10 km/hr), during daylight hours when the sun was well above the horizon, and in winter or spring when cover of green grasses and forbs was minimal. All observers were trained in the identification of PPC and practiced DS and line placement at 2 non-focal sites to perfect techniques before implementation.

For each PPC detected, we gathered the following information: 1) perpendicular distance from the transect line to the center of the PPC to the nearest dm within 0-8 m and typically to the nearest m otherwise (measured with a tape and rangefinder, respectively), 2) height of PPC in cm from the ground to top of the tallest spine (measured with a ruler), 3) width in cm of the PPC or clump (measured with a ruler), 4) the number of pups or small heads, 5) status of plant (e.g., live or dead), 6) whether plants were marked and if so the code, and 7) used a GPS to record UTMs of all individuals.

To assess the influence of various potential covariates of detection probability and quantify environmental conditions along lines, we gathered the following information within 10-m radius plots centered on lines at points placed every 100 m: 1) volume of vegetation between 0-1 m above ground (measured to nearest 10% between 20-80% and 5% otherwise), 2) grass cover (nearest 10% between 20-80% and 5% otherwise), 3) mean understory height of rooted vegetation (e.g., grasses, forbs, sub-shrubs; nearest dm), and 4) size class of dominant soil substrate (1-fine sand with few larger particles, 2-coarser gravel with particles up to about 1 cm diameter, 3–rocky substrate with particles >2 cm diameter). For vegetation volume, we considered vegetation rooted within plots and visually estimated volume assuming 100% volume around plant canopies. For understory height, we visually estimated the mean height of all understory plants excluding cacti and yucca (e.g., those in the lowest vegetation layer) rooted within plots weighted by cover (e.g., larger plants had higher influence than small ones). For grass cover, we considered annual (excluding small basal species such as *Schismus* sp.) and perennial grasses rooted within plots and focused on basal cover. Additionally, for each plot we noted the dominant vegetation community (desert-scrub and semi-desert grassland) and recorded UTM coordinates of all survey points with a GPS. Subsequently, we used the slope and interpolate shape tools in ArcGIS 10.3.1 to estimate the elevation (m) and slope (%) at each point using a 10-m resolution digital elevation model from the National Elevation Dataset available from the U.S. Geological Survey.

<u>*Analyses*</u>: Before analyses we computed the length of all lines by summing inter-point distances and adding the length of any remainders <100 m that were needed to cover plots. To calculate line lengths, we used UTM coordinates and Pythagorean Theorem. To estimate the abundance and density of PPC, we treated each line as a replicate and stratified by site so that estimates were computed for each site. To compute the density of PPC across the entire population of sites, we weighted estimates at each site by the area of sites. Before analyses, we inspected histograms of raw distance data and established bin sizes (e.g., cut-points) of 2.5 m to smooth data, and right truncated 5% of detections to improve model fit (Fig. 2). We used two strategies to estimate density, abundance, detection probability, and other parameters, with use of program Distance version 6.2 (Thomas et al. 2010). First, we fit a simple detection function to data with use of conventional distance sampling procedures. Second, we fit detection functions with covariates with use of multiple-covariates distance sampling to assess the influence of various factors (other than distance) on the scale of detection functions. In both cases, we fit a single detection function to data for all sites combined because sample sizes were insufficient to fit separate functions for each site. As covariates, we considered estimates of vegetation volume, grass cover, understory height, soil substrate size, and slope, which we averaged among points along each line. Estimates of slope were log transformed to minimize the influence of extreme values. To fit detection functions, we first considered each covariate individually, assessed parameter estimates and their standard errors (SE) to identify influential covariates, and then fit additive combinations of some covariates. To select the best approximating model, we ranked models based on Akaike information criterion adjusted for small sample sizes (AIC_c), evaluated the shapes of detection functions, precision of estimates, and goodness-of-fit for highly ranked models, and selected the best overall model from which we made inferences (Thomas et al. 2010). We considered uniform, half-normal (HN), and hazard-rate (HR) detection functions for models without covariates, and HN and HR functions for models with covariates. When fitting HN and HR functions, we considered models with 0-2 cosine, simple polynomial, and hermite adjustment terms. We excluded dead PPC from estimates.

To further understand factors that influenced the observation process during DS, we used multiple linear regression to assess factors that explained variation in detection distances to PPC. Thus, we fit detection distance as a response variable and considered the following potential covariates: mean vegetation volume, grass cover, understory height, soil substrate size, and slope along lines where each PPC was detected, and the height of each PPC. We log transformed some factors including the detection distance to better meet model assumptions.

To quantify the efficacy of DS, we computed percent differences between values from past censuses and estimates of density and abundance from DS (e.g., bias) at the scale of each site and for the overall population. Additionally, we computed Pearson correlation coefficients to quantify the strength of linear relationships between both raw and log-transformed estimates of density from censuses and those from DS. To assess factors that explained bias at the scale of sites, we used linear regression with bias as the response variable, and the following factors as potential explanatory factors: mean vegetation volume, grass cover, understory height, soil substrate size, log slope, elevation, PPC height, and plot area. For the categorical factor census method (e.g., Roller only vs. Roller and repeated monitoring), we used a *t*-test.

To assess environmental factors that explained variation in PPC densities across space, we used our best overall detection function model and post-stratified estimates by line so as to compute densities at the scale of each line. After censoring data from short lines <200-m in length, this resulted in a sample of 76 lines that averaged 476 m (SE = 16) in length along which densities ranged from 0 to 10.5 plants/ha (mean ± SE = 1.5 ± 0.2). We then developed a linear-mixed effect model to assess the influence of various environmental factors on variation in local densities. To develop models, we fit log density as the response variable and considered the following potential explanatory factors, which we generated after eliminating one factor for each correlated ($r \geq 0.66$) pair of factors that we assumed was less descriptive (e.g., understory height because it was correlated with vegetation volume, and elevation which was correlated with substrate size): mean vegetation volume, grass cover (log transformed), soil substrate size, slope (log transformed), and vegetation community. We also considered quadratic terms for all continuous variables. Because the number of potential explanatory factors was high and data to develop candidate models *a priori* was limited, we used stepwise procedures with mixed variable selection and the stepAIC function in the MASS library in R (Venables and Ripley 2002, R Development Core Team 2016) to select explanatory factors. We used Bayesian information criterion (BIC) to guide variable selection because it penalizes model complexity more than AIC_c and reduces chances of overfitting. To adjust for correlations among observations from lines within the same sites, we fit a random intercept for site. All models were fit with the nlme library in R (Pinheiro et al. 2012, R Development Core Team 2016).

RESULTS

Effort and Detections: We recorded 105 live and 15 dead PPC during DS along 36.9 km of line transects across the 11 sites. Linear effort ranged from as low as 866 m at the smallest site (Sopori 3; 4 ha) to 5,745 m at a larger site (Guy Street; 23.8 ha). Across all sites, we measured environmental factors at 476 points, which ranged from as few as 14 points at the smallest site to 70 at a larger site. Effort was similar in the Santa Cruz (n = 6 sites) and Brawley (5) watersheds. Although there were more sites in desert-scrub (7) than semi-desert grassland (4), on average sites in grassland were larger than those in desert-scrub. Elevation ranged from as low as 799 m at Guy Street in the Brawley watershed to 1,092 m at Sycamore Canyon Properties in the Santa Cruz watershed. We completed DS in February, March, November, and December of 2016, and in February 2017.

Model Selection and Detection Probability: We fit 14 candidate models of detection functions that included between 1 and 4 parameters (Table 1). There was strong evidence factors other than distance influenced detection probability (*P*), with little support for a model that included no covariates ($\Delta AIC_c = 4.55$). The top-ranked model included the covariate substrate size, with *P* declining as soils became increasing dominated by large particles ($\beta \pm SE = -0.44 \pm 0.19$). At 10 m from lines, for example, *P* declined from 0.58 in areas with small- to moderate-sized substrates (e.g., 1.6) to 0.35 in areas with moderate- to large-sized substrates (e.g., 2.3; Fig. 3). Although understory vegetation volume (-0.012 ± 0.0066), grass cover (-0.008 ± 0.004), and cactus height (0.045 ± 0.026) influenced *P* in the expected directions when fit independently (Fig. 3), once the effect of substrate size was considered there was little evidence these covariates improved model fit given associated increases in model complexity (Table 1). In contrast, understory vegetation height (-0.0056 ± 0.0071) and slope (-0.091 ± 0.19) had no influence on *P* (Table 1). Regardless of which covariates were included, estimates of density, average *P*, and other parameters were similar at the scale of the overall population (Table 1). In all cases, half-normal key functions with cosine adjustment terms provided the best fit.

Estimates of *P* from the top-ranked model averaged 0.49 (95% CI = 0.42-0.56), with an effective strip width of line transects (e.g., the distance at which $P = 0.5$) of 9.71 m (95% CI = 8.35-11.28; CV = 0.076). At 2 m from lines, *P* averaged 0.96 and declined to 0.92, 0.80, 0.43, and 0.06 at 3, 5, 10, and 20 m from lines, respectively (Fig. 4).

Detection distances to PPC we observed from lines was explained by the height of plants and by mean grass cover ($R^2 = 0.103$), but other factors had little ($p \leq 0.15$ for log slope) or no ($p \geq 0.41$) association with distances after controlling for these factors. On average, detection distances increased by 5.3 ± 1.8% with each 1-cm increase in the height of plants ($p = 0.005$), but decreased 0.82 ± 0.37% ($p = 0.031$) with each 1% increase in mean grass cover. Mean height of PPC detected along lines was 12.2 cm (SE = 0.40) with only 2.5% of individuals ≤2.8 cm and only 10% ≤6.6 cm, indicating few cacti were small.

Density and Abundance: Across the entire population of sites, we estimated a density of 1.465 live individuals/ha, and an abundance of 294 individuals overall. Importantly, precision was relatively high (CV= 0.139; 95% CI in Table 1). These estimates were based on a total sample of 100 individuals after truncating 5% of observations, indicating that we detected approximately one third of all individuals estimated to occur within the boundaries of focal sites. At the scale of individual sites, density estimates ranged from 0.146 to 5.95 individuals/ha and abundance from 3 to 125 individuals, with much lower precision (Table 2).

Efficacy of Distance Sampling: Across the entire population of sites, both density and abundance were fairly well estimated by DS with an overall bias across the population of 11.4%. At the scale of individual sites, however, bias in density estimates ranged from as low as 59.6% underestimation to as high as 64.1% overestimation, with the absolute value of bias as low as

7.3% (Table 2). Density estimates from DS were also highly correlated with estimates based on census efforts on both the untransformed ($r = 0.82$, $p = 0.002$), and especially, logarithmic scales ($r = 0.92$, $p < 0.001$; Fig. 5). Based on an estimated known population size of 332 PPC across all sites, we detected approximately 30% of individuals during DS. With regard to factors that explained bias, there was some evidence bias increased ($\beta \pm$ SE = -1.03 ± 0.65, $p = 0.14$) as mean grass cover increased. Although there was no evidence means differed due to high variability and small sample size ($p = 0.41$), mean bias averaged 14.9 ± 28.5% higher at sites where densities were documented with the Roller method (mean = 18.0% underestimation) than those where the Roller method followed by intensive repeated monitoring was used (mean = 3.0% underestimation).

Factors that Explained Densities: Local densities at the scale of individual lines within sites increased with increasing slope and soil substrate size, and decreased with increasing understory vegetation volume ($p \leq 0.022$) after adjusting for repeated measurements of the same sites (Table 4). Densities decreased by 1.5 ± 0.6% with each 1% increase in grass cover. After accounting for the effects of all three factors, there was no evidence variation in local densities was associated with grass cover ($p = 0.59$) or vegetation communities ($p = 0.21$). Based on the top-ranked model (Table 1), density averaged 1.00 individuals/ha in desert-scrub (CV = 0.224, 95% CI = 0.64-1.56) and 1.84 in semi-desert grassland (CV = 0.235, 95% CI = 1.15-2.93).

DISCUSSION

We tested a new method for estimating abundance of the endangered Pima Pineapple Cactus (PPC) in southern Arizona based on distance-sampling procedures (Buckland et al. 2001). Although distance sampling (DS) is a proven and efficient tool for estimating abundance and detection probability in a broad range of terrestrial and aquatic animal systems (Thomas et al. 2002, 2010), to our knowledge, our efforts represent just its fifth application in a plant system (Buckland et al. 2007, Jensen and Meilby 2012, Kissa and Sheil 2012, Schorr 2013). Overall, results of our efforts were auspicious and suggest DS can provide precise and fairly unbiased estimates of density and other parameters for research and monitoring applications in this and likely other similar systems in the Sonoran Desert. Additionally, we also provided insights into environmental factors that influence detection probability and abundance, which will be useful to managers, policy makers, and researchers in understanding PPC ecology and guiding existing and new survey techniques. Application of DS, however, was far from perfect and thus we suggest some modifications to the procedures used here to improve inference.

Estimates of bias based on what we presumed to be parametric values of densities from census efforts and estimates from DS was fairly low across the population, equaling just 11.4% underestimation overall. At the scale of individual sites, however, bias at some localities was much higher and included both underestimation and overestimation. Importantly, the magnitude of bias seemed relatively consistent across the entire range of densities we considered as indicated by a fairly tight linear relationships and high correlation between values from both procedures. Additionally, the precision of density estimates based on DS was also relatively high (CV =0.139), with 95% confidence intervals that were narrow even despite relatively small sample sizes of 100 individuals. For DS along line transects, a recommended minimum sample size of between 60 and 80 focal objects is recommended to obtain unbiased estimates (Buckland et al. 2001). These results suggest DS can provide relatively accurate estimates of density across the full range of variation in densities we considered (e.g., 0.1-5.5 individuals/ha), and likely, across the natural range of densities that occur in the wild, which are approximately 1 individual/ha on average at large spatial scales across the range of the species (Baker 2013, McDonald 2005).

Two additional factors also provide support for the applicability of DS for PPC population enumeration. Important assumptions of DS include perfect detection of focal objects on the transect line and use of a randomized design to ensure lines are positioned independently of focal objects so that objects are not uniformly distributed with respect to their distances from lines. In cases where individual plants are closely clustered, focal objects may not be distributed uniformly with respect to lines, especially when plots are small (Buckland et al 2007). Frequency histograms of detection distances of PPC had an obvious shoulder and declined relatively monotonically with distance from lines, especially when data were appropriately binned. Such results suggest PPC distribution was sufficiently uniform to eliminate issues imposed by clustering (Buckland et al. 2015). Although PPC were sometimes founds in small groups of several nearby plants, clustering did not seem to impose significant bias, eliminating the need for crossed designs and other approaches for addressing these issues (Buckland et al. 2007). With regard to perfect detection of focal objects on the transect line, there was some evidence negative bias of estimates was due to plants, especially small ones, being hidden by vegetation along lines (see below). Nonetheless, the relative openness of arid environments that provide habitat for this species, the unique silhouette of PPC, and recommendations we summarize below should adequately mitigate this issue. In general, DS is a suitable and efficient method for estimating PPC abundance, and should also be useful for monitoring spatiotemporal changes in distribution and abundance such as has been included in Pima County's Multi-species Conservation Plan (Pima County 2016).

Several factors likely influenced observed bias of estimates, and knowledge of these processes has important implications for understanding our results and guiding future efforts. First, while we assumed values from past censuses represented parametric values, actual population sizes and thus densities were not known exactly. This is because surveys of the five plots we considered in the Brawley watershed were last conducted ≈4 years ago, because new individuals were continuously documented during repeated monitoring, and because PPC populations at these sites declined at an average rate of ≈4.9%/year over 9 years (2003-2012) based on data provided by Baker (2013). Similarly, estimates at sites censused with the Roller method followed by repeated monitoring across time were likely more accurate than those at the four sites where we conducted Roller surveys given greater search effort even despite the temporal issues noted above. Moreover, even by spacing observers at 4-6 m intervals as dictated by the Roller (1996b) method, estimates of detection probability we report here suggest between 4-8% of individuals are likely to be undetected during Roller-type surveys. Such factors likely contributed to bias we observed here and suggests the Roller method does not ensure perfect detection probability. Second, while we attempted to search clumps of dense vegetation near lines for PPC during surveys, there was some evidence bias increased with increasing grass cover. While such patterns are based on small sample sizes, they suggest we failed to detect some PPC on or close to lines, especially when grass cover was high. Finally, distances to observed PPC varied markedly with the height of PPC plants and with grass cover, suggesting we likely missed more small individuals than larger ones, especially in areas with moderate to high grass cover. Together, these factors suggest high likelihoods of negative bias during DS, such as we observed here, and the need for designing surveys to minimize bias associated with these issues.

Some inferences we summarized on habitat relationships are consistent with the known biology of PPC, whereas others varied somewhat. Similar to our results for densities, McPherson (2002) found that occurrence of PPC plants was positively associated with larger-sized soil substrates (e.g., gravel vs. sand). However, Kidder (2015) suggested that at one site higher sand content was associated with larger PPC and more pups. Although McPherson (2002) found the occurrence of PPC plants was associated with moderate levels of herb and woody plant cover, we found that densities declined with increasing grass cover, although few sites had cover that exceeded 15%. Similarly, Kidder (2015) noted that PPC grew in sites that had uniformly high incoming solar radiation (i.e., growing in the open) and equated the open areas where PPC occurred to low levels of competition for soil moisture with other plants. Although McPherson (2002) found occurrence of PPC plants was not associated with any specific landform or slope position, we found local densities within sites increased in areas with high slopes, potentially due to the relationship between steeper slopes and prevalence of larger soil substrates. Differences in the scale of measurements and focal parameter between studies (small-scale, plant-centered plots and occurrence - McPherson 2002; larger area-based scale along lines and density - this study) may explain some differences in observed habitat associations. Regardless, such results suggest the applicability of DS for assessing large-scale habitat relationships. Because our study was not designed specifically to assess habitat relationships, however, results reported here are preliminary.

Although often considered a nuisance parameter, understanding factors that influence detection probability (*P*) is important for designing survey methods because the best techniques have a high and consistent probability of detecting the target species and low sampling error (Thompson et al. 1998). We found that the size of soil substrates had the greatest influence on *P* during DS, with lower *P* associated with larger substrates. A likely explanation for this result is that rockier substrates make cacti more difficult to see by oncoming observers because they break up the unique silhouette of this species. We also found that when considered individually in detection functions, *P* declined with increasing understory vegetation volume and grass cover, and decreasing PPC height. Once the influence of substrate size was considered, however, there was little evidence these covariates significantly influenced the scale of detection functions due likely to small sample sizes. With the addition of more samples in the future, we suspect these factors will improve model fit and accuracy, and thus should be measured as part of DS protocols. This possibility is emphasized by the fact that detection distances increased as the height of plants increased and decreased with grass cover. Although such results suggest we were more likely to miss small PPC during DS, average height of PPC we detected (12.2 cm) was similar to that found by Baker (2013; 10.9 cm) across time, and overall bias was fairly low. These patterns and the fact that DS can provide accurate results despite missing a large portion of focal objects (Buckland et al. 2001) suggest DS is an appropriate technique for estimating PPC abundance at least in populations with typical size distributions. Regardless, the influence of covariates of *P* had relatively small effects on the overall magnitude of density estimates, at least at the population scale, and on average, *P* was high (≈0.50) suggesting DS surveys for PPC are likely to be generally efficient. To our knowledge, this study is the first to explicitly estimate *P* of PPC populations and assess environmental factors associated with variation in *P*.

Recommendations: Despite promising results, several changes in the protocol we used should improve accuracy. First, our estimates of effective strip width and *P* suggest ≈20% of areas between neighboring parallel line transects were not covered. Because considering these areas should improve accuracy, we recommend reducing spacing between lines from 50 to 40 m, and perhaps somewhat closer in areas with dense grass cover. This modification will augment the number of individual cacti detected but may result in a few larger individuals being detected from neighboring lines, which can be addressed by truncation and censoring observations during analyses. Second, more effort should be placed on detecting all individuals, especially smaller ones, on or immediately around the transect line. Such effort could involve somewhat longer search times (e.g., slower walking speeds) and more intensive searches under and around clumps of low vegetation during DS. Third, surveys at sites with steep slopes and dense vegetation along drainage channels were often problematic when lines were not perpendicular to contours. This is because surveying steep slopes and walking through dense vegetation along washes while surveying for cacti was difficult, distracting, and sometimes required slight repositioning of lines to flatter or more open areas (such as Palo Alto where we markedly overestimated densities). To address this issue, we suggest placing lines perpendicular to the slope gradient so that observers walk up and down steep slopes and across washes rather than along contours and drainage channels. Finally, we also recommend that the timing of DS surveys for PPC be focused during periods when herbaceous vegetation cover is likely to be minimal and when grasses are not green. In our region, this time period is often between November and June unless fall and winter rains have been substantial. While additional field study and simulations across a gradient of contexts are needed to better understand the efficacy of DS for PPC, results obtained here together with the above recommendations offer promising opportunities. Finally, because DS can easily be completed within the context of Roller-type surveys without significant increases in effort, we recommend DS be integrated into existing protocols to facilitate additional study.

ACKNOWLEDGEMENTS

We thank Robert Villa for assisting with DS surveys, and Sherry and Michael Mann for completing Roller surveys of four plots in the Santa Cruz Valley in southern Arizona. Marc Baker graciously provided shape files and information from his long-term monitoring efforts in the Altar Valley, and Scott Hart of Westland Resources provided information for Sycamore Canyon Properties. We thank Pat and John King of the King Anvil Ranch for permission to cross their land to access plots. We thank Julie Crawford of U.S. Fish and Wildlife Service for providing information on the Pima Pineapple Cactus, and Deborah Atkinson of Arizona Department of Agriculture for administrative support.

LITERATURE CITED

Anderson, D.R., Burnham, K.P., Lubow, B.C., Thomas, L., Corn, P.S., Medica, P.A. and Marlow, R.W. 2001. Field trials of line transect methods applied to estimation of desert tortoise abundance. Journal of Wildlife Management 65:583-597.

Baker, M. 2013. A demographic study of *Coryphantha robustispina* ssp. *robustispina* - Progress report for the 2012 field season and final report for the study. November 7, 2013. 66 pp.

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, UK.

Buckland, S.T., Borchers, D.L., Johnston, A., Henrys, P.A. and Marques, T.A. 2007. Line transect methods for plant surveys. Biometrics, 63:989-998.

Buckland, S.T., Rexstad, E.A., Marques, T.A. and Oedekoven, C.S. 2015. Distance sampling: methods and applications. Springer, New York, NY.

Flesch, A.D., González Sánchez, C., Valenzuela Amarillas J. 2016. Abundance and habitat relationships of breeding birds in the Sky Islands and adjacent Sierra Madre Occidental of northwest Mexico. Journal of Field Ornithology 87:176-195.

Jensen, A. and H. Meilby. 2012. Assessing the population status of a tree species using distance sampling: *Aquilaria crassna* (Thymelaeaceae) in Northern Laos. International Journal of Forestry Research. http://dx.doi.org/10.1155/2012/265831

Kidder, A.L. 2015. Ecohydrological conditions associated with the distribution and phenology of the Pima pineapple cactus (*Coryphantha scheeri* var. *robustispina*). Master of Science Thesis, School of Natural Resource, University of Arizona, Tucson, Arizona.

Kissa, D.O. and Sheil, D. 2012. Visual detection based distance sampling offers efficient density estimation for distinctive low abundance tropical forest tree species in complex terrain. Forest Ecology and Management, 263:114-121.

McDonald, C.J. 2005. Conservation of the rare Pima pineapple cactus (*Coryphantha scheeri* var. *robustispina*): Recruitment after fires and pollination in the Altar Valley of southern Arizona. Master of Science Thesis, School of Natural Resource, University of Arizona, Tucson, Arizona.

McPherson, G.R. 2002. Relationship of ecological variables in the field with the presence of Pima pineapple cactus. Final Report Submitted to United States Fish and Wildlife Service, Tucson, Arizona, April 2002.

Pima County. 2016. Multi-species Conservation Plan for Pima County, Arizona: Final. Submitted to the Arizona Ecological Services office of the U.S. Fish and Wildlife Service, Tucson, Arizona.

Pinheiro, J., Bates, D., Debroy, S., Sarkar, D. and R Development Core Team. 2012. nlme: Linear and Nonlinear Mixed Effects Models. R

Powell, B.F. 2015. Pima pineapple cactus surveys on select parcels of the Marley Ranch, Pima County, Arizona. Unpublished report by the Pima County Office of Sustainability and Conservation, Tucson, Arizona.

R Development Core Team. 2016. R: a language and environment for statistical computing. Version 3.3.1.R Foundation for Statistical Computing, Vienna, Austria.

Roller, P.S. 1996a. Distribution, growth, and reproduction of Pima pineapple cactus (*Coryphantha scheeri* Kuntz var. *robustispina* Schott). Master of Science Thesis, School of Natural Resource, University of Arizona, Tucson, Arizona.

Roller, P. 1996b. Pima pineapple cactus recommended survey protocol: Three tier survey methods. Accessed on May 30, 2014 from: https://www.fws.gov/southwest/es/arizona/Documents/SpeciesDocs/PimaPineappleCactus/ppc_survey_protocol.pdf

Rosenstock, S.S., Anderson, D.R., Giesen, K.M., Leukering, T. and Carter M.F. 2002. Landbird counting techniques: Current practices and an alternative. Auk 119:46-53.

Routson, R. 2003. Demographic Study of *Coryphantha scheeri* var. *robustispina*. Report submitted to the Arizona Sonora Desert Museum; Tucson, Arizona.

Schorr, R.A. 2013. Using distance sampling to estimate density and abundance of *Saussurea weberi* Hultén (Weber's Saw-Wort). The Southwestern Naturalist 58:378-383.

Schmalzel, R. 2000. Growth and age-structure of a clonal cactus, *Coryphantha scheeri* var. *robustispina*, and notes on its performance with respect to soil age and to banner-tailed kangaroo rat mounds. West San Pedro Pasture, King Anvil Ranch. Arizona Department of Agriculture Final Report.

Thomas, L., Buckland, S.T., Burnham, K.P., Anderson, D.R., Laake, J.L., Borchers, D.L. and Strindberg, S. 2002. Distance sampling. Environmetrics 1:544-552.

Thomas, L., Buckland, S.T., Burnham, K.P., Anderson, D.R., Laake, J.L., Borchers, D.L. and Strindberg, S. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology 47:5-14.

Thompson, W.L., White, G.C. and Gowan, C. 1998. Monitoring vertebrate populations. Academic Press, San Diego, California.

USFWS [U.S. Fish and Wildlife Service]. 2007. Five-year review: Pima pineapple cactus. Albuquerque, New Mexico.

USFWS [U.S. Fish and Wildlife Service]. 2017. Draft recovery plan for *Coryphantha scheeri* var. *robustispina* (Pima pineapple cactus). U.S. Fish and Wildlife Service, Southwest Region, Tucson, Arizona.

Venables, W.N. and B.D. Ripley. 2002. Modern applied statistics with S, fourth edition. Springer, New York, NY.

WestLand Resources, Inc. 2004. A preliminary population estimate of Pima pineapple cactus (*Coryphantha robustispina*) in south-central Arizona. Tucson, Arizona.

WestLand Resources, Inc. 2017. Pima pineapple cactus monitoring at Sycamore Canyon transplant areas, 2016. Report submitted to Mr. Jamie Argueta, Sycamore Canyon

Table 1: Candidate models of detection functions we considered when estimating the density of the Pima Pineapple Cactus based in distance sampling at 11 sites in south-central Arizona, 2016-17. *K* denotes the number of model parameters, *D* is estimated density (no. of live individuals/ha), CV is the coefficient of variation, *N* is total abundance or population size, LCL and UCL are lower and upper 95% confidence intervals, ESW is effective strip width, and *P* is average detection probability. Estimates are from program Distance (version 6.2; Thomas et al. 2010), based on a sample of 105 cacti detected (5% of observations truncated), and all models are based on half normal key functions with cosine adjustments.

	Model Selection		Density				Abundance			Detection	
Covariates	*K*	ΔAICc	*D*	*D* CV	*D* LCL	*D* UCL	*N*	*N* LCL	*N* UCL	ESW	*P*
Substrate Size	2	0.00	1.465	0.139	1.109	1.937	294	222	388	9.71	0.485
Substrate Size + Grass Cover	3	0.12	1.484	0.140	1.120	1.965	297	224	394	9.59	0.479
Cactus Height + Substrate Size + Grass Cover	4	1.44	1.493	0.140	1.126	1.979	299	226	397	9.53	0.476
Cactus Height + Substrate Size	3	1.52	1.472	0.139	1.113	1.948	295	223	390	9.66	0.483
Substrate Size + Vegetation Volume 0-1 m	3	1.86	1.467	0.139	1.109	1.941	294	222	389	9.69	0.485
Cactus Height + Grass Cover	3	2.06	1.466	0.139	1.108	1.941	294	222	389	9.70	0.485
Vegetation Volume 0-1 m	2	2.98	1.440	0.138	1.091	1.901	289	219	381	9.88	0.494
Grass Cover	2	3.17	1.439	0.138	1.090	1.899	288	218	381	9.89	0.494
Cactus Height	2	3.21	1.434	0.138	1.087	1.893	287	218	379	9.92	0.496
Cactus Height + Grass Cover + Veg. Volume 0-1 m	4	3.85	1.470	0.140	1.109	1.947	294	222	390	9.68	0.484
None {CDS model}	1	4.55	1.406	0.141	1.059	1.866	282	212	374	10.12	0.506
Grass Cover + Vegetation Volume 0-1 m	3	4.87	1.443	0.138	1.092	1.907	289	219	382	9.85	0.493
Understory Height	2	5.88	1.412	0.137	1.072	1.860	283	215	373	10.07	0.504
Slope (log)	2	6.41	1.408	0.136	1.069	1.853	282	214	371	10.11	0.505

Table 2: Comparison of estimates of density (*D*) and abundance (*N*) of the Pima Pineapple Cactus based in distance sampling at 11 sites and all sites combined in south-central Arizona, 2016-17. Census results are based on the Roller (1996) method and the Roller method followed by intensive repeated monitoring over time, and completed within 0 to 4 years of distance sampling effort. Bias denotes the % difference between census results and estimates from distance sampling. Population estimates are area-weighted averages. Distance-based estimates are from program Distance (version 6.2; Thomas et al. 2010), based on a sample of 105 cacti detected (5% of observations truncated), and based on a half normal key function with cosine adjustment.

		Census Results				Distance Sampling Results						Bias (%)	
Site	Plot Area (ha)	*D*	*N*	Source	Method	*D*	*N*	CV	No. Observed	Effort (m)	No. of Lines	*D*	*N*
Anvil*	18.3	0.219	4	Baker	Monitoring	0.146	3	1.004	1	3,525	7	-33.1	25.0
Canoa	23.4	2.35	55	This study	Roller	1.07	25	0.301	10	4,825	8	-54.6	-54.5
Guy Street	23.8	0.252	6	Baker	Monitoring	0.179	4	0.733	2	5,745	11	-28.8	-33.3
Mendoza*	24.2	2.85	69	Baker	Monitoring	1.30	32	0.287	13	5,133	10	-54.2	-53.6
Palo Alto	24.6	3.26	80	Baker	Monitoring	5.02	125	0.231	38	3,902	10	56.0	56.3
Sopori 1	7.4	1.62	12	This study	Roller	2.18	16	0.292	4	947	3	34.7	33.3
Sopori 2	8.0	3.86	31	This study	Roller	1.56	12	0.452	6	1,985	6	-59.6	-61.3
Sopori 3	4.0	5.53	22	This study	Roller	5.95	24	0.302	10	866	4	7.6	9.1
Stagecoach	31.6	0.222	7	Baker	Monitoring	0.363	11	0.469	3	4,252	8	64.1	57.1
Sycamore 1	16.7	1.86	31	Westland	Monitoring	1.72	29	0.342	9	2,696	9	-7.3	-6.5
Sycamore 2	18.4	0.817	15	Westland	Monitoring	0.686	13	0.505	4	3,003	7	-16.0	-13.3
All Sites	200.4	1.66	332	---	---	1.47	294	0.139	100	36,878	83	-11.4	-11.4

*Distance sampling transects covered only portions of original plots containing the majority of the cactus population, with plot area and densities adjusted from those reported in Baker (2013). The Anvil plot contained 4 of 5 known plants and the Mendoza plot contained 69 of 71 known plants.

Table 3: Comparison of geographic, topographic, and vegetation factors at 11 sites where we implemented distance sampling for the Pima Pineapple Cactus in south-central Arizona, 2016-17. Means and standard errors (SE; or range) are based on sample sizes (*n*) noted for each site, which are based on measurements at points (elevation, slope), within 10 m of points (vegetation factors), or at the site scale (region, dominant vegetation community).

				Elevation (m)		Slope (%)		Substrate Size		Grass Cover (%)		Vegetation Volume 0-1 m (%)		Understory Height (cm)	
Site	Region	Vegetation Community	*n*	Mean	Range	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Anvil	Brawley	Grassland	42	829	8	1.5	0.1	1.0	0.00	13.7	2.5	13.5	2.2	14.1	1.6
Canoa	Santa Cruz	Grassland	56	934	15	3.7	0.2	2.1	0.11	64.3	4.0	53.2	2.2	63.5	2.5
Guy Street	Brawley	Desert-scrub	70	802	7	1.8	0.1	1.1	0.03	1.1	0.3	14.2	1.2	10.3	0.8
Mendoza	Brawley	Grassland	66	978	18	5.4	0.3	1.9	0.08	38.1	2.3	25.5	1.9	28.6	1.2
Palo Alto	Brawley	Grassland	48	890	20	8.0	1.1	1.7	0.11	10.4	1.6	22.4	2.4	27.9	2.1
Sopori 1	Santa Cruz	Desert-scrub	14	991	9	6.1	0.7	2.2	0.15	3.8	1.0	25.0	5.1	14.4	1.5
Sopori 2	Santa Cruz	Desert-scrub	30	992	11	7.9	1.3	2.2	0.16	3.2	0.9	24.2	2.8	16.6	1.7
Sopori 3	Santa Cruz	Desert-scrub	14	985	11	6.6	0.9	2.0	0.00	2.0	1.0	15.5	2.7	11.5	1.5
Stagecoach	Brawley	Desert-scrub	58	1,027	21	3.2	0.1	1.1	0.05	3.9	0.5	14.3	1.1	13.9	1.1
Sycamore 1	Santa Cruz	Desert-scrub	36	1,083	14	3.2	0.1	2.7	0.09	3.0	1.0	31.9	3.0	29.3	3.9
Sycamore 2	Santa Cruz	Desert-scrub	42	1,003	15	2.8	0.1	2.7	0.11	3.1	1.0	34.7	3.1	16.1	2.3

Table 4: Factors that explained variation in local densities (log no./ha) of the Pima Pineapple Cactus along 76 line transects at 11 sites in southern Arizona, 2016-17. Parameter estimates and standard errors (SE) are from a linear mixed-effects model in which site was fit as a random intercept (σ^2 = 0.03 intercept; 0.27 residual) and estimates of local density derived from distance sampling was fit as the response variable. Non-significant factors are not included in this table but noted in the text.

Factor	Estimate	SE	|*t*|	*p*
Intercept	-0.62	0.31	1.98	0.052
Vegetation Volume 0-1 m (%)	-0.015	0.006	2.36	0.022
Slope (log %)	0.65	0.19	3.45	0.001
Substrate Size (rank)	0.36	0.16	2.28	0.026

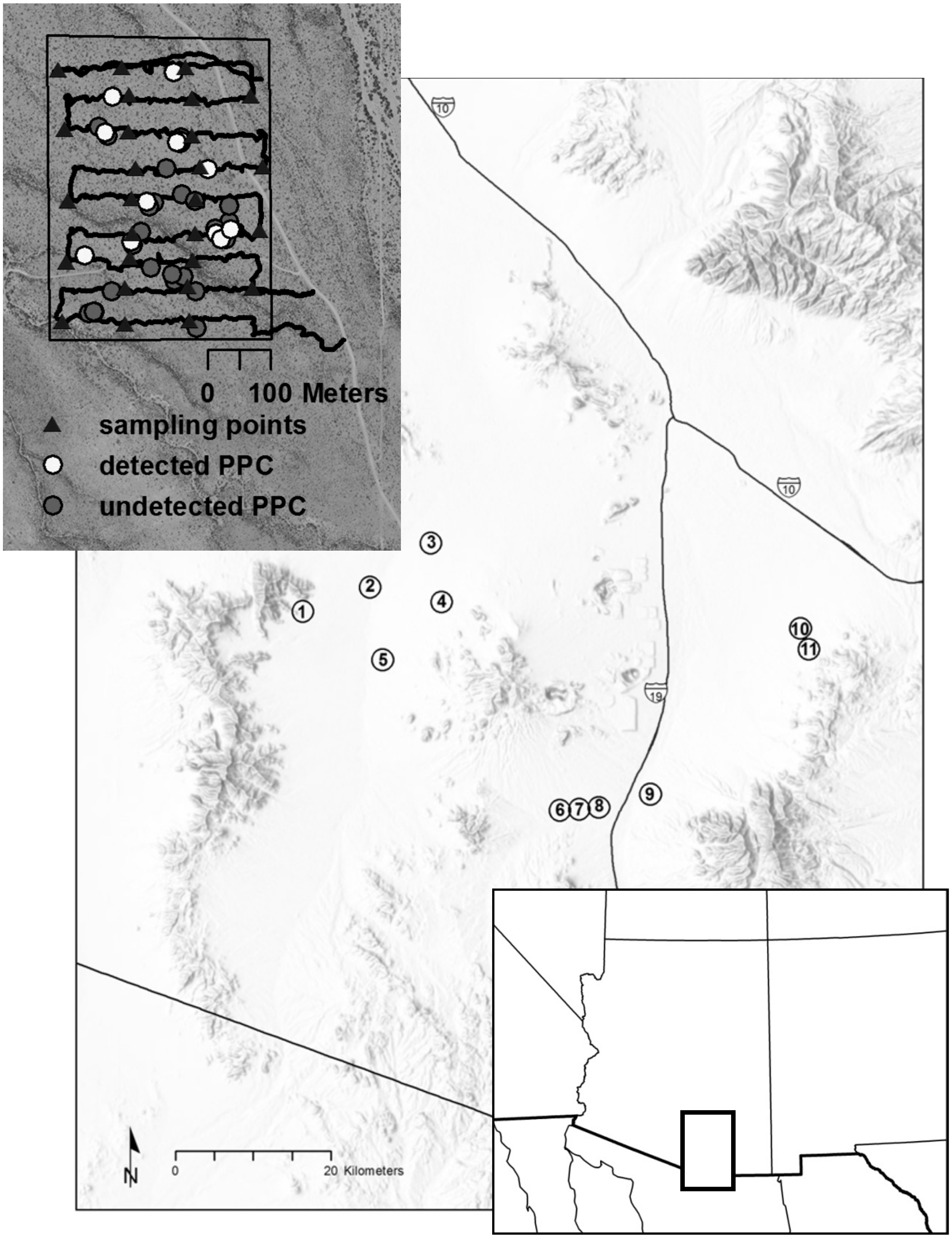


Figure 1: Location of 11 sites where we estimated densities and detection probability of the Pima Pineapple Cactus with use of distance-sampling methods in south-central Arizona, 2016-17. Top inset figure shows the arrangement of line transects and cacti detected and not detected at site number 11 and the sampling points located at 100-m intervals at which we sampled environmental attributes, and lower inset shows the location of the study area with reference to state and national boundaries. Sites are as follows: 1) Mendoza, 2) Anvil, 3) Guy Street, 4) Stagecoach, 5) Palo Alto, 6-8) Sopori 1-3, 9) Canoa, 10-11) Sycamore 1-2.

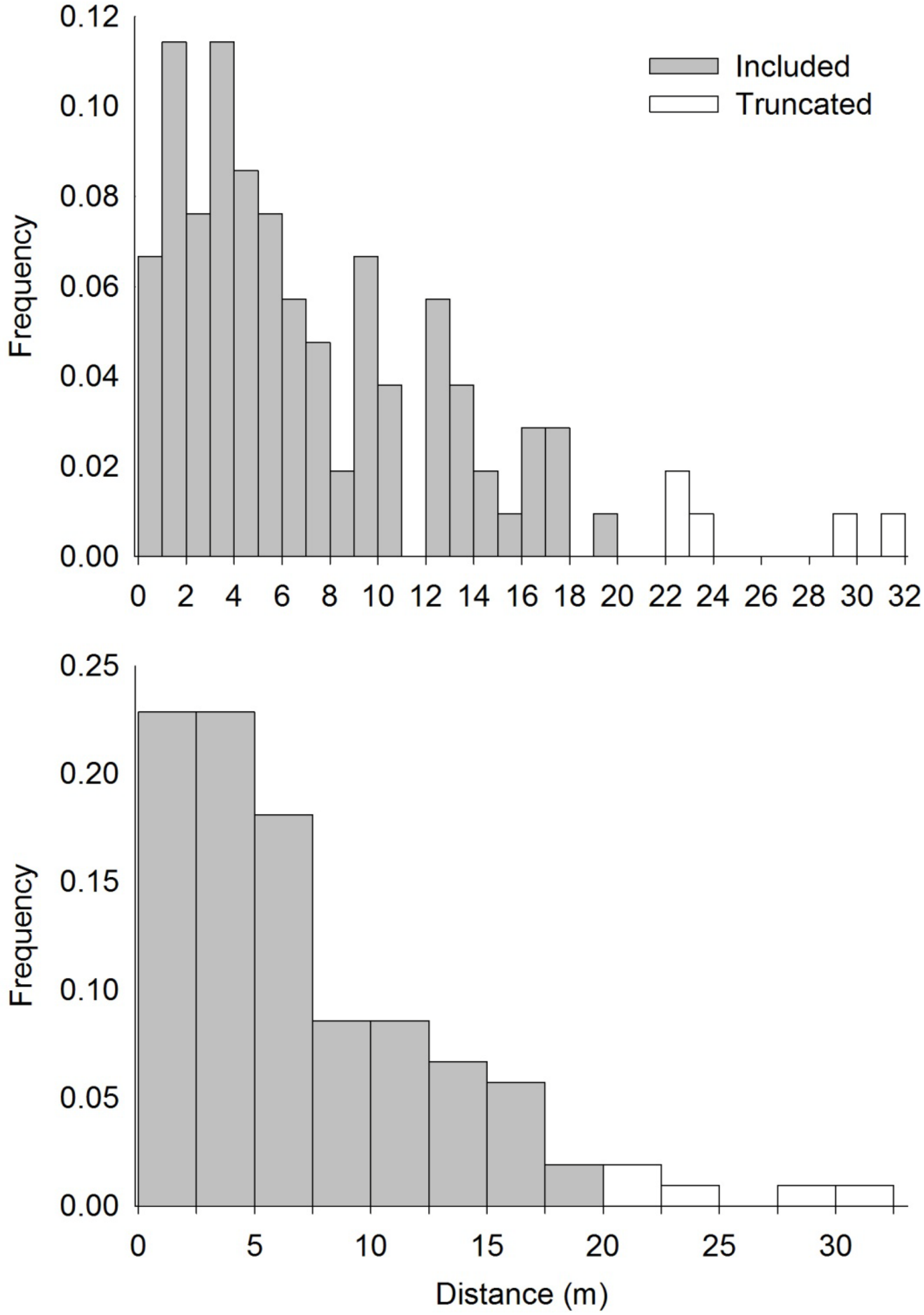


Figure 2: Frequency histograms of detection distances of 105 Pima Pineapple Cacti observed during line-transect surveys in southern Arizona, 2016-17. Top figure shows raw frequencies within 1-m bins and lower figure shows frequencies within the 2.5 m bins used when modeling detection functions. Open bars at distances >20 m represent 5% of observations we truncated when fitting detection functions.

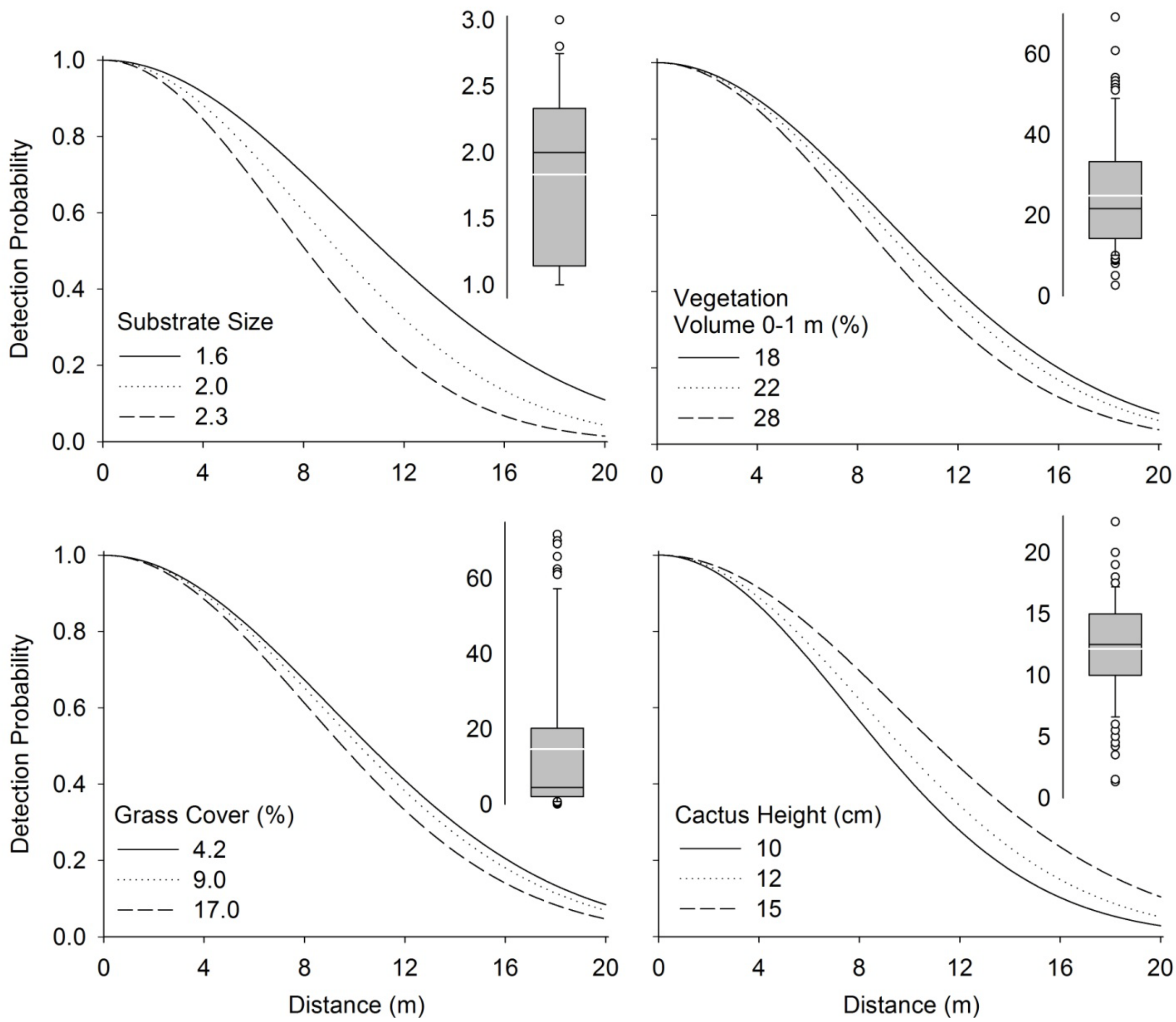


Figure 3: Influence of four covariates on detection probability of the Pima Pineapple Cactus from distance sampling along line transect at 11 sites in south-central Arizona, 2016-17. Estimates are based on multiple covariates distance sampling and half normal key functions with cosine adjustments in which each covariate was fit individually. Estimates are shown at covariate levels equaled to the lower, middle (e.g., median), and upper quartiles, which are indicated by the bottom, black line, and top of inset box plots that show the distribution of each covariate (white lines are means). Model selection criteria for each model are provided in Table 1.

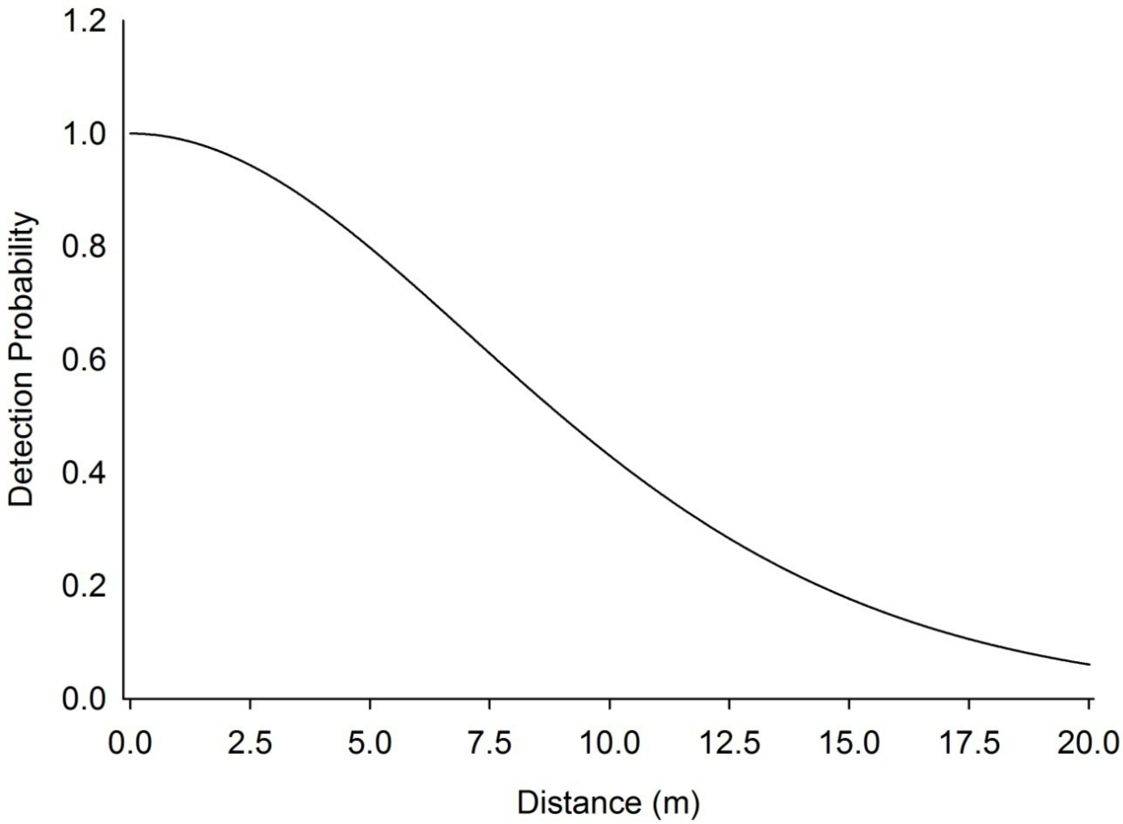
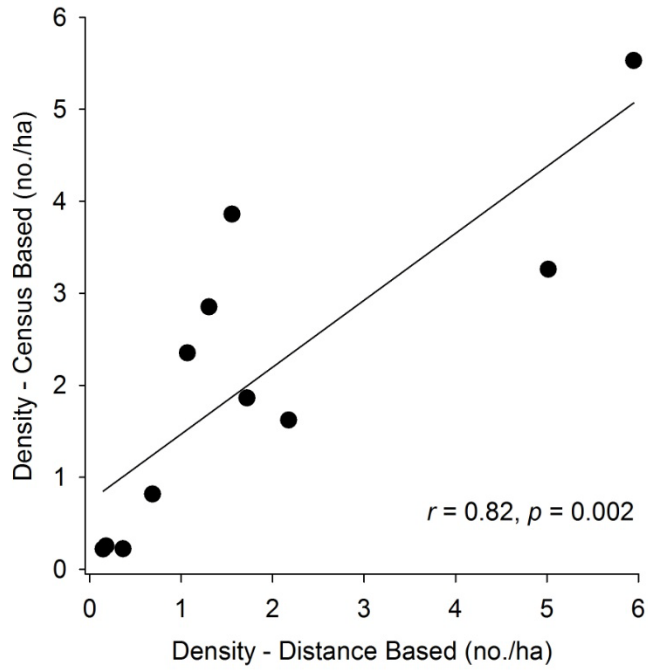


Figure 4: Top-ranked detection function model for the Pima Pineapple Cactus based on distance sampling along line transect at 11 sites in south-central Arizona, 2016-17. Estimates are based on 100 observations, multiple covariates distance sampling, and a half normal key function with cosine adjustments in which substrate size was fit as a covariate. The plotted function is the average detection function conditional on the observed covariates.



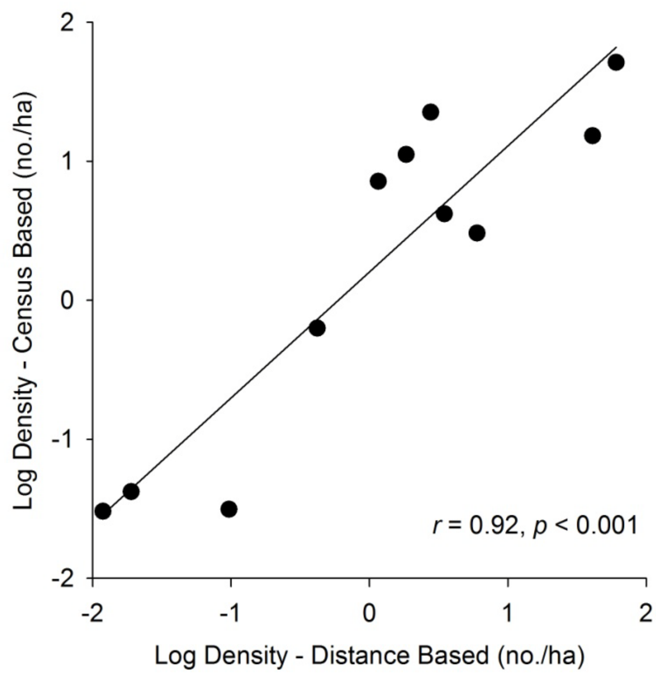


Figure 5: Linear associations between estimates of density (no./ha; top) and log density (bottom) of the Pima Pineapple Cactus at 11 sites in south-central Arizona, 2016-17. Estimates from distance sampling are based on 100 observations, multiple covariates distance sampling, and a half normal key function with cosine adjustments in which substrate size was fit as a covariate. Estimates from censuses were based on the Roller (1996b) method often combined with repeated monitoring and searches across time. Pearson correlation coefficients (*r*) are shown on figures.