Application of distance sampling for assessing abundance and habitat relationships of a rare Sonoran Desert cactus

Aaron D. Flesch • Ian W. Murray • Jeffrey M. Gicklhorn • Brian F. Powell

Aaron D. Flesch
School of Natural Resources and the Environment, University of Arizona, The Desert Laboratory on Tumamoc Hill, 1675 W. Anklam Road, Tucson, AZ 85745, USA
email: flesch@email.arizona.edu, phone: 520-629-9455, ORC ID: 0000-0003-3434-0778

Ian W. Murray, Jeffrey M. Gicklhorn
Pima County Office of Sustainability and Conservation
201 N. Stone Ave, Tucson, AZ 85701, USA

Brian F. Powell
Pima County Natural Resources, Parks and Recreation
3500 W River Rd, Tucson, AZ 85741, USA

Abstract

Accurate abundance estimates of plant populations are fundamental to numerous ecological questions and for conservation. Estimating population parameters for rare or cryptic plant species, however, can be challenging and thus developing and testing new methods is useful. We assessed the efficacy of distance sampling for estimating abundance and habitat associations of the endangered Pima pineapple cactus (*Coryphantha scheeri* var. *robustispina*), a rare plant in the Sonoran Desert of southwestern North America that has traditionally been surveyed with census-based methods. Distance sampling (DS) involves measuring distances between focal objects and samples of lines or points, and modeling detection functions that adjust estimates for variation in detection probability (*P*). Although often used in animal systems, DS remains largely untested for plants. We encountered 105 live individuals along 36.9 km of transects in 11 study plots placed across much of the geographic range of the species, and estimated an average density of 1.47 individuals/ha (CV = 0.139). Compared to values from intensive censuses, density estimates from DS were underestimated by only 2.3% on average and highly correlated on the untransformed ($r = 0.84$) and logarithmic ($r = 0.93$) scales. Estimates of *P* averaged 0.49 and declined as soils became increasingly dominated by larger soil substrates, and somewhat with increasing vegetation volume and decreasing cactus height. Local densities increased with increasing slope and soil substrate size and decreased with increasing vegetation volume ($p \leq 0.024$). Combined with careful survey design, DS offers an efficient method for estimating population parameters for uncommon and cryptic plants.

Keywords Abundance estimation • Detection probability • Distance sampling • Habitat • Pima pineapple cactus • Population size • *Coryphantha scheeri* var. *robustispina*

Introduction

Estimating the abundance and habitat associations of plant populations is fundamental to a broad range of ecological questions and for guiding conservation and management. For uncommon and cryptic plants, however, accurate estimates of population parameters can be costly to obtain, and thus developing new more efficient methods is useful. Although many species of plants are readily detectable in the field, probability of detecting individuals that are present and available for sampling is rarely perfect and can vary with species' traits, and environmental and survey conditions (Chen et al., 2009, 2013; Garrard et al. 2013; Junaedi et al. 2018). Understanding factors that influence the detection process can help guide survey and sampling designs, and explicitly modeling these factors can improve the accuracy of inferences (Buckland et al. 2001; Chen et al. 2009; Dénes et al. 2015). Plant species that occur as scattered individuals have often been surveyed with plot-less or point-based techniques that are sometimes referred to as distance methods (Cottam 1947; Cottam and Curtis 1956; Mueller-Dombois and Ellenberg 1974; Elzinga et al. 1998). For these species, such techniques are thought to be faster and more flexible than plot-based methods, but can be challenging to implement in field settings when individuals are rare or cryptic (Elzinga et al. 1998; Ducey 2018).

Distance sampling (DS) is a survey technique similar to—but distinct from—traditional distance methods in plant ecology. This approach involves measuring distances to focal objects from sets of lines or points, and modeling a detection function that quantifies the decline in detection probability with increasing distance from observers, and adjusts abundance estimates for variation in detection probability (Buckland et al. 2001; Thomas et al. 2010). In addition to observed distances, other covariates of detection probability such as individual (e.g., plant size), spatial, and temporal factors can be incorporated into detection function. Hence, DS offers excellent flexibility and can be tailored to specific traits of focal populations and their environment (Marques et al. 2007). Application of DS has proven highly effective for estimating abundance and habitat relationships of wildlife, and been used across a broad range of geographic regions and taxa (Thomas et al. 2002, 2010; Anderson et al. 2001; Rosenstock et al. 2002; Hounsome et al. 2005; Flesch et al. 2016). Although commonly applied to wildlife, DS remains largely untested for plants. To date, DS has been applied to few plant systems (e.g., Buckland et al. 2007; Crase et al. 2010; Kissa and Sheil 2012; Schorr 2013), and its efficacy has not been tested based on parametric values of plant abundance or used to evaluate plant-habitat relationships.

We assessed the efficacy of DS for estimating abundance of a rare plant in an arid environment. As a case study, we considered the Pima pineapple cactus (*Coryphantha scheeri* var. *robustispina* Britton and Rose, Cactaceae; hereafter "PPC"), an endangered species in the Sonoran Desert of southwestern North America. Like many species of concern, the PPC is often surveyed for compliance with federal law and to address conservation and recovery objectives. The recommended survey method for this species, however, calls for a complete census of all individuals in a given focal area, which is time intensive and costly (Roller 1996a; USFWS 2007, 2018). In this and other similar contexts, survey methods based on sampling theory should be capable of accurately estimating population size, distribution, and other parameters with greater efficiency across larger areas.

We compared estimates of population size and densities of the PPC derived from DS to values from intensive recent censuses, and assessed the magnitude of estimation bias and factors that explain bias. Moreover, we evaluated factors that influence detectability during DS and the resulting implications for survey design. Finally, we assessed plant-habitat relationships by modeling variation in local densities and various environmental factors such as vegetation structure and soil substrate size.

Materials and Methods

Study system

The PPC is distributed narrowly in the eastern Sonoran Desert of south-central Arizona and adjacent Sonora, Mexico (Baker and Butterworth 2013). In Arizona, it occurs near the ecotone of Sonoran desert-scrub and semi-desert grasslands in the Altar and Santa Cruz valleys (Fig. 1; USFWS 2018). Individuals are small (≤46 cm in height), hemispherical succulents with singular or clumped stems covered by 2-3-cm long rounded projections (USFWS 2018). Sonoran desert-scrub is dominated by small leguminous trees such as velvet mesquite (*Prosopis velutina*) and paloverde (*Parkinsonia* sp.), shrubs such as creosote (*Larrea tridentata*) and bursage (*Ambrosia* sp.), and various cacti, grasses, and forbs (Turner and Brown 1982). Semi-desert grassland is dominated by open woodlands of velvet mesquite and various grasses and sub-shrubs such as burroweed (*Isocoma tenuisecta*) and snakeweed (*Gutierrezia sarothrae*; Brown and Makings 2014).

In response to threats from urban development, invasion of non-native grasses, wildfire, climate change, and other stressors, the PPC was listed as endangered in 1993 (USFWS 2007; Thomas et al. 2017). The recommended survey protocol for this species attempts to census all individuals in a given focal area (Roller 1996a). While this approach is useful for compliance with U.S. federal law, it is inefficient for other objectives and may be based on unrealistic assumptions of perfect detection probability. Such issues are especially relevant because individuals are small, widely spaced, and sometimes concealed by dense vegetation, which augments the chances some individuals are undetected during surveys. To guide conservation and recovery, efficient survey techniques for estimating abundance across large areas are needed together with data on habitat associations.

Design

We implemented DS at sites across the northern range of the PPC (Fig. 1), along broad natural gradients in densities, and assessed estimation bias by comparing abundance estimates from DS with values from intensive censuses. We selected study plots where either: 1) repeated censuses and monitoring had enumerated PPC abundances within ≈1 year, or 2) presence of PPC was known but abundances were unknown and thus needed to be measured prior to DS. For criteria one, we selected five plots in the Altar Valley and two plots in the Santa Cruz Valley where long-term monitoring began in the late 1990s and 2004, respectively (Fig. 1; see Appendix A for details). For criteria two, we selected four additional plots where observers different from those that implemented DS completed intensive censuses in the same year. To census plots, multiple observers walked parallel lines 4-6 m apart and exhaustively searched for cacti until plots were completely covered (Roller 1996a). Within long-term study plots (criteria one), all known cacti were monitored and plots surveyed at 1-4 year intervals with new individuals added to results. Thus, assuming accuracy of past censuses and population closure, populations within plots were completely enumerated within ≈1 year of DS.

We systematically placed parallel lines 50-m apart across plots and began DS from a random point on plot boundaries. To guide survey design, we used estimates from preliminary PPC surveys along 37 km of lines in similar environments (B. Powell, unpubl. data), which found an effective strip half-width (distance from the line at which the number of focal objects missed equals the number detected beyond that distance but within the truncation distance; Buckland et al. 2015) of 8-13 m and maximum detection distance of 25 m. To assess environmental conditions along lines, we measured various environmental factors, which are described below, around points placed every 100 m (Fig. 2).

Surveys and measurements

For stationary objects, DS has two assumptions to ensure accurate estimation: 1) perfect detection of focal objects on survey lines (or at points), and 2) accurate measurements of distances between lines and objects. Moreover, lines should be placed independently of focal objects so objects are uniformly distributed with respect to distances from lines (Buckland et al. 2015).

During DS, teams of two observers slowly walked lines. One observer focused on and immediately around lines while another observer walked short serpentine paths within ≈0-6 m of lines scanning lines and surrounding areas (Fig. 2). Observers inspected vegetation clumps near lines to ensure cacti on lines were detected and looked behind them for cacti to check for individuals obstructed from oncoming directions. Because PPC sometimes occur in small groups 10-30 m apart, before leaving lines to measure detected cacti, observers scanned areas for additional individuals. All surveys were during daylight hours when the sun was well above the horizon and in winter and early spring when cover of green grasses and forbs was low.

We recorded the following data for each PPC: 1) perpendicular distance from transect line to center of cactus, 2) height of cactus (cm) from ground to top of tallest spine, 3) width (cm) of cactus, 4) number of pups or stems, 5) status of cactus (live or dead), and 6) location based on GPS coordinates. We used measuring tapes to estimate distances to the nearest dm within 0-8 m of lines, or laser rangefinders to the nearest m beyond 8 m, and used tapes to measure cacti dimensions to the nearest 0.5 cm.

To assess the influence of potential covariates of detection probability and quantify local environmental conditions, we estimated environmental features within 10-m radius plots centered on points placed every 100 m along lines. We estimated: 1) vegetation volume from 0-1 m above ground, 2) percent grass cover, 3) mean understory height of vegetation, and 4) size class of dominant soil substrate. Volume and cover were visually estimated to the nearest 10% for values between 20-80% and nearest 5% otherwise. When measuring volume, we considered vegetation rooted within plots and assumed 100% volume around plant canopies. For grass cover, we considered annual and perennial grasses rooted within plots but excluded small prostrate species (e.g., *Schismus* sp.). For understory height, we visually estimated the mean height of understory plants rooted within plots, which included grasses, forbs, and sub-shrubs but excluded succulents. For soil substrate size, we considered three size classes; 1 for fine sand with few larger particles, 2 for coarser gravel with particles up to about 1 cm diameter, and 3 for rocky substrate with particles >2 cm diameter. For each plot we noted dominant vegetation community as Sonoran desert-scrub or semi-desert grassland. Subsequently, we used the slope and interpolate shape tools in ArcGIS 10.5.1 (ESRI 2017) to estimate elevation (m) and slope (%) at each point based on a 3-m resolution digital elevation model.

Analyses

To estimate abundance and density, we treated lines as replicates and stratified by plot to facilitate estimates at both scales, and then weighted by plot areas to estimate overall population size across all plots. Before analyses, we selected bin sizes of 2.5 m after assessing histograms of distance data, and right truncated 5% of observations. Binning can improve model fit by effectively smoothing data whereas truncation constrains the tails of distributions, which often include little information but require complex adjustment terms to model that are rarely biologically justified (Buckland et al. 2001; Thomas et al. 2010).

We used two strategies to estimate density, population size, and detection probability of live individuals. First, we used conventional distance sampling to fit a detection function to all data. Second, we used multiple-covariates distance sampling to fit detection functions that included each covariate individually and various additive combinations of covariates. As covariates, we considered vegetation volume, grass cover, understory height, soil substrate size, and slope averaged among points along each line. To minimize influence of extreme values, we log transformed slope. To select the best approximating model, we ranked models by Akaike information criteria corrected for small sample sizes (AIC_c), evaluated shapes of detection functions, precision of estimates, and goodness-of-fit among competitive 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 ≤2 cosine, simple polynomial, and hermite adjustment terms. We used program Distance version 6.2 for all calculations (Thomas et al. 2010). Although we sampled without replacement and detected a relatively large proportion of the focal population, finite population correction factors were not applied but may be appropriate here despite limited influence on estimates of precision (see Buckland et al. 2001:87).

To assess the efficacy of DS, we computed bias as the percent difference between values from censuses and estimates from DS within each plot and for the overall population. To quantify the strength of linear association between census values and estimates, we computed Pearson correlation coefficients on both the raw and log-transformed scales. To assess factors that explained bias at the scale of plots, we used linear regression with bias as a response variable, and considered mean vegetation volume, grass cover, understory height, soil substrate size, log slope, elevation, PPC height, and plot area as potential explanatory variables. Finally to compare effort needed to complete DS vs. censuses, we calculated the total effort spent DS on a per ha basis and compared to estimates for censuses based on data from Roller (1996a), which indicates a minimum of 2.3 person hrs are required per ha.

To assess environmental factors that explained spatial variation in local densities among lines, we fit linear-mixed effect models. To develop models, we fit log density as a response variable and considered the following potential explanatory factors: mean vegetation volume, grass cover (log transformed), soil substrate size, log slope, vegetation community, and quadratic terms for all continuous factors. Understory height was not considered because it was correlated with vegetation volume, nor was elevation considered because it was correlated with substrate size ($r \geq 0.65$). 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 from the MASS library in R (Venables and Ripley 2002; R Core Team 2016) to guide model selection. We fit a random intercept for plot to adjust for correlations among observations from lines within the same plots, and fit models with the nlme library in R (Pinheiro et al. 2012, R Core Team 2016). Data from short lines (<200 m) needed to cover irregularly shaped plots were censored because they contained too few (0-2) environmental sampling points to adequately describe local conditions.

Results

Effort and detections

We recorded 105 live and 15 dead PPC during DS along 36.9 km of transects (n = 81 lines, mean ± SE = 455 ± 17.9 m in length) across the 11 plots. Distances between lines and cacti averaged 7.3 ± 0.6 m (range 0-31 m) with 75% of observations within 10.3 m, and 2.5 m binning and 5% truncation effectively smoothed data (Fig. 3). Although more plots were in desert-scrub, total plot area was similar in both vegetation communities (Table S1).

Detection probability

We fit 14 candidate models of detection functions (Tables 1 and S2). Model selection provided strong evidence that factors in addition to distance influenced detection probability (P), as indicated by little support for a model without covariates (ΔAIC_c = 4.55). The top-ranked model included the covariate substrate size, with P declining as soils became increasing dominated by large particles (β ± SE = -0.44 ± 0.19). At 10 m from lines, for example, P declined from 0.58 in areas with small- to moderate-sized substrates to 0.35 in areas with moderate- to large-sized substrates (Fig. 4). 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. 4), there was little evidence these factors improved models once substrate size was considered (Table 1). Understory vegetation height (-0.0056 ± 0.0071) and slope (-0.091 ± 0.19) had no influence on P. 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 half-width of 9.7 m (95% CI=8.4-11.3; 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. 3).

Abundance and bias estimation

Across the entire population of plots, we estimated a density of 1.47 live individuals/ha, and abundance of 294 individuals overall. Precision of estimates was fairly high (CV= 0.139; Table 2). At the plot scale, estimates of density (0.17-5.95 individuals/ha) and abundance (3-125 individuals) ranged widely, with much lower precision (Table 2). Estimates of population size suggest we detected approximately 34% of all individuals during DS.

Across all plots, DS provided relatively unbiased estimates of both density and abundance, with estimation bias averaging only -2.3% overall. At the scale of individual plots, however, estimates of bias were higher (Table 2). Density estimates from DS were also highly correlated with census values on both the untransformed (r = 0.84, p = 0.002), and especially, logarithmic scales (r = 0.93, p < 0.001; Fig. 5). Bias decreased (e.g., changed from over to underestimation) as substrate size ($\beta \pm$ SE = -66.6 ± 27.7, p = 0.040) and understory vegetation volume (-3.3 ± 1.5, p = 0.050) increased. On average, DS took 0.60 ± 0.06 person hrs per ha to implement across plots (range = 0.35-1.05) with effort increasing linearly with plot-specific PPC densities ($\beta \pm$ SE = 0.094 ± 0.022, p = 0.0019). Thus, we estimate censuses would take a minimum of 4.2 ± 0.4 times more effort to complete on average across the range of PPC densities we considered.

Plant-habitat relationships

We considered a total sample of 76 lines averaging 476 m (SE = 16) in length with densities ranging from 0 to 10.5 plants/ha (mean ± SE = 1.5 ± 0.2). Local densities increased with increasing soil substrate size and slope, and decreased with increasing understory vegetation volume (Table 3). There was also some evidence local densities were greater in semi-desert grasslands than in desert-scrub, with densities averaging 38.3 ± 17.2% greater in grasslands after controlling for other factors. Local densities did not vary with grass cover ($p = 0.59$) after considering factors in the best approximating model.

Discussion

We validated a rarely used method for estimating the abundance and density of plant populations. Our study, focused on the endangered Pima pineapple cactus (PPC) in the Sonoran Desert, indicates that distance sampling (DS) can efficiently provide accurate estimates of abundance, and insights into factors that explain local variation in densities and detection probability. Combined with results from a small number of past applications of DS in plant systems, our results indicate that DS is an efficient tool in this and other similar systems, and useful for guiding management and survey design. Distance sampling has been used successfully to assess abundance and detection probability (P) in animal systems, often at much larger spatial scales than that considered here (Thomas et al. 2002, 2010; Buckland et al. 2015; Roberts et al. 2016). To our knowledge, however, this study represents just its eighth application in a plant system (Marsden and Pilgrim 2003; Buckland et al. 2007; Crase et al. 2010; Jensen and Meilby 2012; Kissa and Sheil 2012; Schorr 2013; Phama et al. 2014; Junaedi et al. 2018), and is the first to compare estimates from DS with what we assumed were parametric values of abundances from intensive censuses.

Bias of abundance estimates from DS was very low across the sampled population, averaging just 2.3% underestimation. Importantly, magnitude of bias seemed consistent across the entire range of abundances we considered, except perhaps at lower extremes, suggesting DS performs well across broad spatial variation in abundance. Such results conform generally to studies in animal populations (e.g., Focardi et al. 2005) where DS has accurately captured major declines in densities despite lower precision at low densities, but to our knowledge, no comparable examples exist for plant populations. At very low densities, small differences in estimates and parametric values can have marked effects on bias. In these and other cases, stratification and fitting stratum as a factor-type covariate should enhance precision by explicitly modeling spatial differences in abundance (Buckland et al. 2015).

Precision of estimates from DS was also fairly high at population scales (CV= 0.139) even despite modest sample sizes of 105 individuals along 81 lines. For DS along lines, a recommended minimum of 60-80 focal objects (or clusters) are recommended for unbiased estimation (Buckland et al. 2001). These results and the broad range of natural variation in PPC densities we considered (0.1-5.5 individuals/ha), suggests DS can yield precise abundance estimates in a range of contexts. Interestingly, our estimate of PPC density (1.47 individuals/ha) was higher than range-wide estimates of $\approx$1 individual/ha (Baker 2013; McDonald 2005) likely because we worked in areas where PPC was known to occur.

Important assumptions of DS along lines include perfect detection of focal objects on transect lines, accurate distance measurements, and designs that ensure lines are positioned independently of focal objects. If individual plants are closely clustered, distributions may not be sufficiently uniform with respect to lines, especially in small plots (Buckland et al 2007). In our study, frequency histograms of detection distances declined monotonically with increasing distance from lines, especially after data were smoothed by binning. Such patterns suggest PPC distribution is sufficiently uniform to eliminate issues imposed by clustering (Buckland et al.

2007, 2015), even though plants sometimes occurred in small groups of individuals 5-20 m apart. In other systems, more extreme clustering may require crossed designs or more complex approaches (see Buckland et al. 2007), or cluster-based estimation where numbers of individuals in clusters is used as detection covariate (Thomas et al. 2010).

With regard to assumptions of perfect detectability of focal objects on transect lines, there was some evidence small plants obstructed by dense vegetation and rocky substrates contributed to underestimation. Nonetheless, the relative openness of arid environments and unique silhouette of PPC should adequately mitigate these issues, especially when combined with recommendations described below. In plant systems such as ours where individual plants are small, often cryptic, scattered over large areas, and thus easily overlooked, DS should be an efficient method for estimating spatiotemporal variation in abundance.

Several factors likely contributed to observed estimation bias. First, while we assumed numbers from past censuses represented parametric values of population sizes, actual abundances were not known exactly. Because plots were censused within ≈1 year of DS, the closure assumption (e.g., no recruitment or mortality) was likely violated. Data from seven plots that were intensively monitored over time indicate abundance declined by an average rate of 7.3% per year between 2002 and 2017 (Appendix A). Thus, mortalities occurring after recent censuses, but before DS, could explain some observed bias. Individuals undetected during censuses were occasionally found during subsequent monitoring, because even by spacing observers 4-6 m apart during census efforts (Roller 1996a), estimates of P obtained here suggest 4-8% of individuals are likely to be missed. Finally, although probably a very minor source of bias in our study, ensuring distances are measured precisely and perpendicularly to lines will reduce bias (Marshall et al. 2008).

Detectability-corrected estimates of densities from DS are often used to understand wildlife-habitat relationships (Blank 2013; Miller et al. 2013; Flesch et al. 2016; Roberts et al. 2016), but have not been applied to plants. Past accounts of habitat relationships of PPC often matched our inferences but sometimes varied. Similar to our results for densities, McPherson (2002) found positive associations between PPC occurrence and larger soil substrates (gravel vs. sand), whereas Kidder (2015) suggested sandy soils were associated with larger cacti. The main pollinator of the PPC is a solitary bee (*Diadasia rinconis*) that nests in well-drained areas of bare ground and forages over large areas (Ordway 1987; McDonald 2005; USFWS 2018). Thus, despite local associations with rockier substrates, this species may require a diversity of soils at larger scales. We found that PPC densities declined with increasing grass cover matching observed associations with open areas (Kidder 2015), but contrasting with one study that showed associations between occurrence and moderate levels of herbaceous and woody vegetation cover (McPherson 2002). Moreover, we found local densities increased with slope, which may be partially due to the fact that areas with higher slopes also often have larger soil substrates. These patterns contrast lack of observed associations between occurrence and specific landforms or slope positions (McPherson 2002), and may not be biologically important given limited variation in slope across plots we considered. Importantly, differences in the scales of measurement and focal parameters among studies may explain differences in observed habitat associations. Regardless, our results illustrate the application of DS for assessing plant-habitat relationships.

Understanding factors that influence P is useful for guiding survey design because optimal survey techniques have a high and consistent probability of detecting the target species and low sampling error (Thompson et al. 1998; Williams et al. 2002). Although few studies assess factors that influence detectability of plants, traits such as color, flowering time, leaf size, height, and observer ability can influence the detection process (Chen et al. 2013; Garrard et al. 2013; Junaedi et al. 2018). We found that detectability during DS was explained by variation in soil substrate size, with lower detectability on rockier soils. Rocky substrates likely make cacti more difficult to see by distorting their unique silhouette. There was also some evidence detectability declined with decreasing plant height and increasing understory vegetation volume and grass cover, patterns we suspect are biologically significant and would have been stronger with greater sample sizes. These patterns and the efficacy of DS where significant proportions of focal populations are undetected, support the application of DS for estimating abundance of rare or cryptic plants.

Recommendations

Despite promising results, various design considerations and small modifications to the protocol used here could further improve applications of DS in plant systems. Though our results suggest that only approximately 33% of cacti within plots were detected, one of the strengths of DS is that it allows robust estimates of density and population size even when a majority of focal objects are not detected during surveys (Anderson et al. 2001; Buckland et al. 2001). Precision of estimates from DS, however, are influenced by the absolute number of observations and thus sampling strategies that yield large sample sizes are optimal. In systems where focal plants may be obstructed by either live or dead vegetation, however, more effort on and immediately around lines should improve accuracy. Such effort can be fostered by reducing walking speeds and searching clumps of low vegetation along lines. In our system, surveys on steep slopes and dense vegetation along drainages were often difficult when lines were parallel to elevation contours. Although positioning lines perpendicular to contours will ameliorate these issues (Schorr 2013), to foster unbiased estimates, investigators should ensure lines are placed parallel to any existing density gradients of focal objects, which can be assessed during pilot efforts (Buckland et al. 2015). Finally, timing surveys when focal plants are most detectable (e.g., flowering, in leaf, etc.), associated vegetation is dormant or least obstructive, and measuring factors thought to influence the detection process should further improve efforts.

Although we focused on small plots to help foster comparisons with known values of abundance, DS is most powerful when applied at much larger spatial scales (e.g., Flesch et al. 2016; Roberts et al. 2016), where it can produce reliable estimates provided key assumptions are met and a sufficient number of focal objects are detected. Thus, future studies of plant abundance and detection probability based on DS can be framed much more extensively than the largely intensive focus used here. While our results provide strong support for the application of DS, additional field work and simulations across a range of contexts and efforts with populations where abundances are known, will be useful for guiding future efforts.

Acknowledgements

We thank S. Mann, M. Garcia and R. Villa for field assistance, M. Baker and S. Hart for long-term monitoring data, and P. and J. King for access. J. Crawford of USFWS, D. Atkinson of Arizona Department of Agriculture (ADA), and S. McMahon of University of Arizona provided helpful comments. This study was funded by ADA through USFWS Cooperative Endangered Species Conservation Fund Grant Program (project no. Segment 19, 2015-2017-04).

Compliance with ethical standards

Conflict of interest

No potential conflict of interest was reported by the authors.

Author contributions

BFP largely conceived of the study that was designed by ADF, IWM, and BFP. Data were gathered by ADF, IWM, S. Mann, M. Garcia and R. Villa. ADF completed the analyses with assistance from IWM and JMG. The first draft of the manuscript was written by ADF with all authors commenting on and contributing to subsequent drafts. All authors read and approved the final manuscript.

References

Anderson DR, Burnham KP, Lubow BC, et al. (2001) Field trials of line transect methods applied to estimation of desert tortoise abundance. *J Wildl Manage* **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. Report for USDI Bureau of Reclamation, Grant No. R08AP32230, 66 pp.

Baker M, Butterworth C (2013). Geographic distribution and taxonomic circumscription of populations within *Coryphantha* section *Robustispina* (*Cactaceae*). *Am J Bot* **100**:984-997.

Blank PJ (2013) Northern bobwhite response to conservation reserve program habitat and landscape attributes. *J Wildl Manage* **77**:68-74.

Brown DE, Makings E (2014) A guide to grasslands of North America. *Des Plants* **29**:1-160.

Buckland ST, Anderson DR, Burnham KP, et al. (2001) *Introduction to distance sampling: estimating abundance of biological populations*. Oxford, UK: Oxford University Press.

Buckland ST, Borchers DL, Johnston A, et al. (2007) Line transect methods for plant surveys. *Biometrics* **63**:989-998.

Buckland ST, Rexstad EA, Marques TA, et al. (2015) *Distance sampling: methods and applications*. New York, NY: Springer.

Chen G, Kery M, Zhang J, et al. (2009) Factors affecting detection probability in plant distribution studies. *J Ecol* **97**:1383-1389.

Chen G, Kery M, Plattner M, et al. (2013) Imperfect detection is the rule rather than the exception in plant distribution studies. *J Ecol* **101**:183-191.

Cottam G (1947) A point method for making rapid surveys of woodlands. *Bull Ecol Soc Am* 28:60.

Cottam G, Curtis JT (1956) The use of distance measures in phytosociological sampling. *Ecology* **37**:451-460

Crase B, Duguid A, Pengart RN, Jakamarra PW, Angal LP, Pengart MS, Kemarr AW (2010) Distribution and conservation status of the giant sweet potato, a rare Aboriginal food plant from Central Australia. *North Territ Nat* **22**:17-30.

Dénes, FV, Silveira, LF, Beissinger SR (2015) Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. Methods in Ecology and Evolution 6:543-556.

Ducey MJ (2018) Design-unbiased point-to-object sampling on lines, with applications to areal sampling. *Eur J For Res* **137**:367-383.

Elzinga CL, Salzer DW, Willoughby JW (1998) *Measuring and monitoring plant populations.* U.S. Department of the Interior, Bureau of Land Management, Technical Reference 1730–1, Denver, Colorado.

ESRI (Environmental Systems Research Institute) (2017) *ArcGIS Release 10.5.1.* Redlands, CA.

Flesch AD, 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. *J Field Ornithol* **87**:176-195.

Focardi S, Montanaro P, Isotti R, et al. (2005) Distance sampling effectively monitored a declining population of Italian roe deer *Capreolus capreolus italicus*. *Oryx* **39**:421-428.

Garrard GE, McCarthy MA, Williams NS, et al. (2013) A general model of detectability using species traits. *Methods Ecol Evol* **4**:45-52.

Hounsome TD, Young RP, Davison J, et al. (2005) An evaluation of distance sampling to estimate badger (*Meles meles*) abundance. *J Zool* **266**:81-87.

Jensen A, Meilby H (2012) Assessing the population status of a tree species using distance sampling: *Aquilaria crassna* (Thymelaeaceae) in Northern Laos. *Int J For Res*, doi:10.1155/2012/265831.

Junaedi DI, McCarthy MA, Guillera-Arroita G, et al. (2018) Traits influence detection of exotic plant species in tropical forests. *PloS One*, e0202254.

Kidder AL (2015) Ecohydrological conditions associated with the distribution and phenology of the Pima pineapple cactus (*Coryphantha scheeri* var. *robustispina*). *Masters thesis*. University of Arizona, Arizona.

Kissa DO, Sheil D (2012) Visual detection based distance sampling offers efficient density estimation for distinctive low abundance tropical forest tree species in complex terrain. *Forest Ecol Manag* **263**:114-121.

Marsden SJ, Pilgrim JD (2003) Diversity and abundance of fruiting trees in primary forest, selectively logged forest, and gardens on New Britain, Papua New Guinea. *Tropical Biodiversity* **8**:15-29.

Marques TA, Thomas L, Fancy SG, et al. (2007) Improving estimates of bird density using multiple-covariate distance sampling. *Auk* **124**:1229-1243.

Marshall AR, Lovett JC, White PC (2008) Selection of line-transect methods for estimating the density of group-living animals: lessons from the primates. *Am J Primatol* **70**:452-462.

McDonald CJ (2005) Conservation of the rare Pima pineapple cactus (*Coryphantha scheeri* var. *robustispina*): Recruitment after fires and pollination in the Altar Valley of southern Arizona. *Masters thesis*. University of Arizona, Arizona.

McPherson GR (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.

Miller DL, Burt ML, Rexstad EA, et al. (2013) Spatial models for distance sampling data: recent developments and future directions. *Methods Ecol Evol* **4** :1001-1010.

Mueller-Dombois D, Ellenberg H (1974) *Aims and methods of vegetation ecology*. London, U.K: John Wiley and Sons.

Ordway E (1987) The life history of *Diadasia rinconis* Cockerell (Hymenoptera: Anthophoridae). *J Kansas Entomol Soc* **60**:15-24.

Pinheiro J, Bates D, Debroy S, et al. (2012) *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-140.

Phama JO, Panagos MD, Myburgh WJ, et al. (2014) The population status of the endangered endemic plant *Aloe peglerae*: Area of occupancy, population structure, and past population trends. S Afr J Bot **93**:247-251.

Powell, BF (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 Core Team (2016) *R: a language and environment for statistical computing.* Version 3.3.1.R Vienna, Austria: Foundation for Statistical Computing.

Roberts JJ, Best BD, Mannocci L, et al. (2016) Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. *Sci Rep* **6**:22615.

Roller PS (1996a) *Pima pineapple cactus recommended survey protocol: Three tier survey methods.* Accessed on May 15, 2019 from: https://www.fws.gov/southwest/es/arizona/Documents/SpeciesDocs/PimaPineappleCactus/ppc_survey_protocol.pdf

Roller PS (1996b) Distribution, growth, and reproduction of Pima pineapple cactus (*Coryphantha scheeri* Kuntz var. *robustispina* Schott). *Masters thesis*. University of Arizona, Arizona.

Rosenstock SS, Anderson DR, Giesen KM, et al. (2002) Landbird counting techniques: Current practices and an alternative. *Auk* **119**:46-53.

Schorr RA (2013) Using distance sampling to estimate density and abundance of *Saussurea weberi* Hultén (Weber's Saw-Wort). *Southwest Nat* **58**:378-383.

Thomas L, Buckland ST, Burnham KP, et al. (2002) Distance sampling. *Environmetrics* **1**:544-552.

Thomas L, Buckland ST, Burnham KP, et al. (2010) Distance software: design and analysis of distance sampling surveys for estimating population size. *J Appl Ecol* **47**:5-14.

Thomas KA, Jarchow CJ, Crawford JA (2017) Survival of the Endangered Pima Pineapple Cactus: Does clearing before prescribed fire alter survival postfire? *Southwest Nat* **62**:200-206.

Thompson WL, White GC, Gowan C (1998) *Monitoring vertebrate populations*. San Diego, California: Academic Press.

Turner RM, DE Brown (1982) Sonoran desertscrub. In Brown, D (ed). Biotic Communities of the American Southwest – United States and Mexico. *Desert Plants* **4**:181-221.

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

USFWS (U.S. Fish and Wildlife Service) (2018) Recovery plan for *Coryphantha scheeri* var. *robustispina* (Pima pineapple cactus). U.S. Fish and Wildlife Service, Southwest Region, Tucson, Arizona.

Venables WN, Ripley BD (2002) *Modern applied statistics with S, fourth edition*. New York, NY: Springer.

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

Table 1 Detection function models fit to estimate abundance of Pima pineapple cactus in southern Arizona, 2016-17. All models are half normal key functions with cosine adjustments; *K* is the number of parameters and *D* is estimated density (live individuals/ha).

Covariates	*K*	ΔAIC_c	*D*
Substrate Size	2	0.00	1.465
Substrate Size + Grass Cover	3	0.12	1.484
Cactus Height + Substrate Size + Grass Cover	4	1.44	1.493
Cactus Height + Substrate Size	3	1.52	1.472
Substrate Size + Vegetation Volume 0-1 m	3	1.86	1.467
Cactus Height + Grass Cover	3	2.06	1.466
Vegetation Volume 0-1 m	2	2.98	1.440
Grass Cover	2	3.17	1.439
Cactus Height	2	3.21	1.434
Cactus Height + Grass Cover + Veg. Volume 0-1 m	4	3.85	1.470
None {CDS model}	1	4.55	1.406
Grass Cover + Vegetation Volume 0-1 m	3	4.87	1.443
Understory Height	2	5.88	1.412
Slope (log)	2	6.41	1.408

Table 2 Comparison of estimates of density (D) and abundance (N) of the Pima pineapple cactus based on distance sampling at plot-specific and population scales in southern Arizona, 2016-17. Census values are from data in Appendix A. Bias denotes % differences between census values and estimates from distance sampling. Population-scales estimates are weighted by plot area.

		Census Values		Distance Sampling Estimates						Bias (%)	
Site	Plot Area (ha)	Density (no./ha)	Abundance	Density (no./ha)	Abundance	CV	No. Observed	Effort (m)	No. of Lines	Density	Abundance
Anvil	18.3	0.055	1	0.146	3	1.004	1	3,525	7	167.4	200.0
Canoa	23.4	2.35	55	1.07	25	0.301	10	4,825	8	-54.6	-54.5
Guy Street	23.8	0.168	4	0.179	4	0.733	2	5,745	11	6.9	0.0
Mendoza	24.2	1.86	45	1.30	32	0.287	13	5,133	10	-29.8	-28.9
Palo Alto	24.6	3.18	78	5.08	125	0.231	38	3,902	10	60.0	60.3
Sopori-1	7.4	1.62	12	2.18	16	0.292	4	947	3	34.7	33.3
Sopori-2	8.0	3.86	31	1.56	12	0.452	6	1,985	6	-59.6	-61.3
Sopori-3	4.0	5.53	22	5.95	24	0.302	10	866	4	7.6	9.1
Stagecoach	31.6	0.222	7	0.363	11	0.469	3	4,252	8	64.1	57.1
Sycamore-1	16.7	1.86	31	1.72	29	0.342	9	2,696	9	-7.3	-6.5
Sycamore-2	18.4	0.817	15	0.686	13	0.505	4	3,003	7	-16.0	-13.3
All Sites	**200.4**	**1.50**	**301**	**1.47**	**294**	**0.139**	**100**	**36,878**	**83**	**-2.3**	**-2.3**

Table 3 Factors that explained variation in local densities (log no./ha) of Pima pineapple cactus along 76 lines in southern Arizona, 2016-17. Parameter estimates and standard errors (SE) are from a linear mixed-effects model with plot fit as a random intercept ($\sigma^2 = 0.021$ intercept; 0.268 residual).

Factor	Estimate	SE	$\|t\|$	*p*
Intercept	-0.90	0.33	2.73	0.008
Vegetation Volume 0-1 m (%)	-0.021	0.006	3.19	0.002
Slope (log %)	0.69	0.21	3.34	0.001
Substrate Size (rank)	0.40	0.17	2.31	0.024
Semi-desert Grassland	0.38	0.17	2.22	0.053

Figure Captions

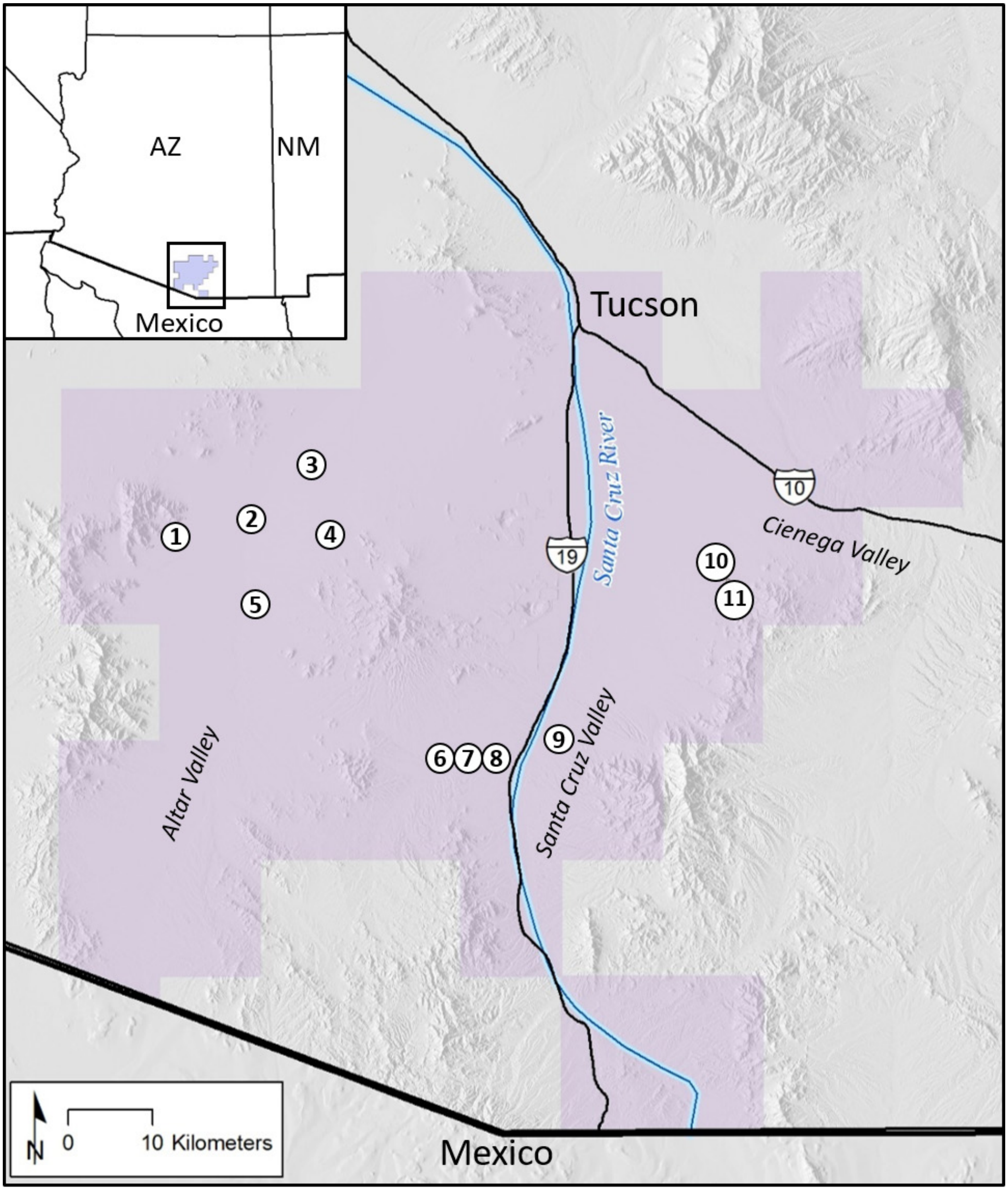
Fig. 1 Plot locations and approximate geographic range (purple) of the Pima pineapple cactus in southern Arizona. Plots are: 1) Mendoza, 2) Anvil, 3) Guy Street, 4) Stagecoach, 5) Palo Alto 6-8) Sopori 1-3, 9) Canoa, and 10-11) Sycamore 1-2.

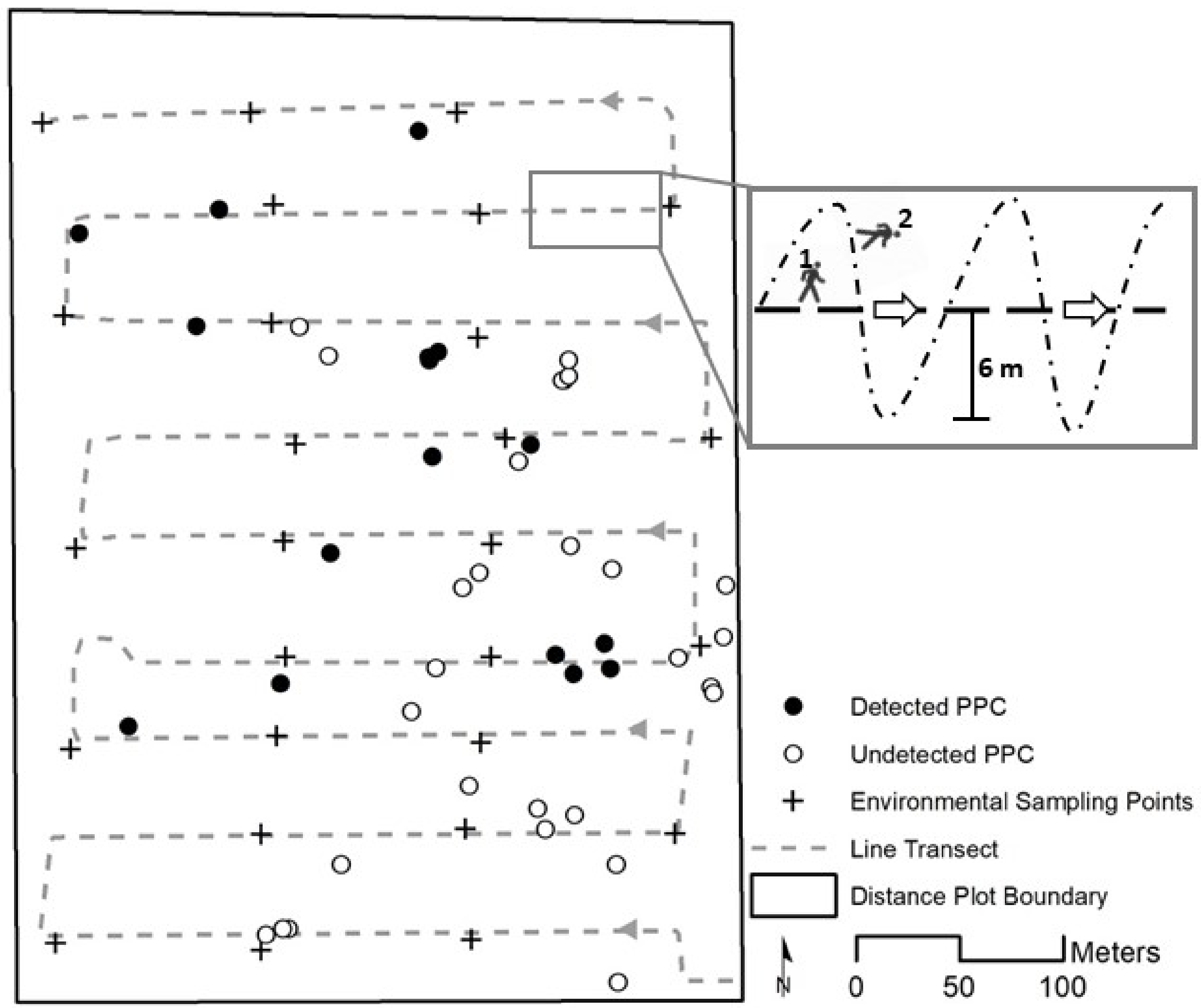
Fig. 2 Arrangement of transect lines (---) and environmental sampling points (+) used to distance sample Pima pineapple cactus (PPC) in southern Arizona, 2016-17. Inset shows sampling strategy by each of two surveyors along lines, with one surveyor focused on center lines, while a second surveyor walked a sinuous path within 6 m of lines.

Fig. 3 Detection distances to 105 Pima pineapple cacti observed during distance sampling in southern Arizona, 2016-17, and resulting detection function model. Frequency histograms of observations in 1- (top) and 2.5-m bins used for modeling (middle) are shown. Detections at distances >20 m shown as open bars were truncated before model fitting. Bottom figure of top-ranked detection function model is the average function conditioned on the covariates.

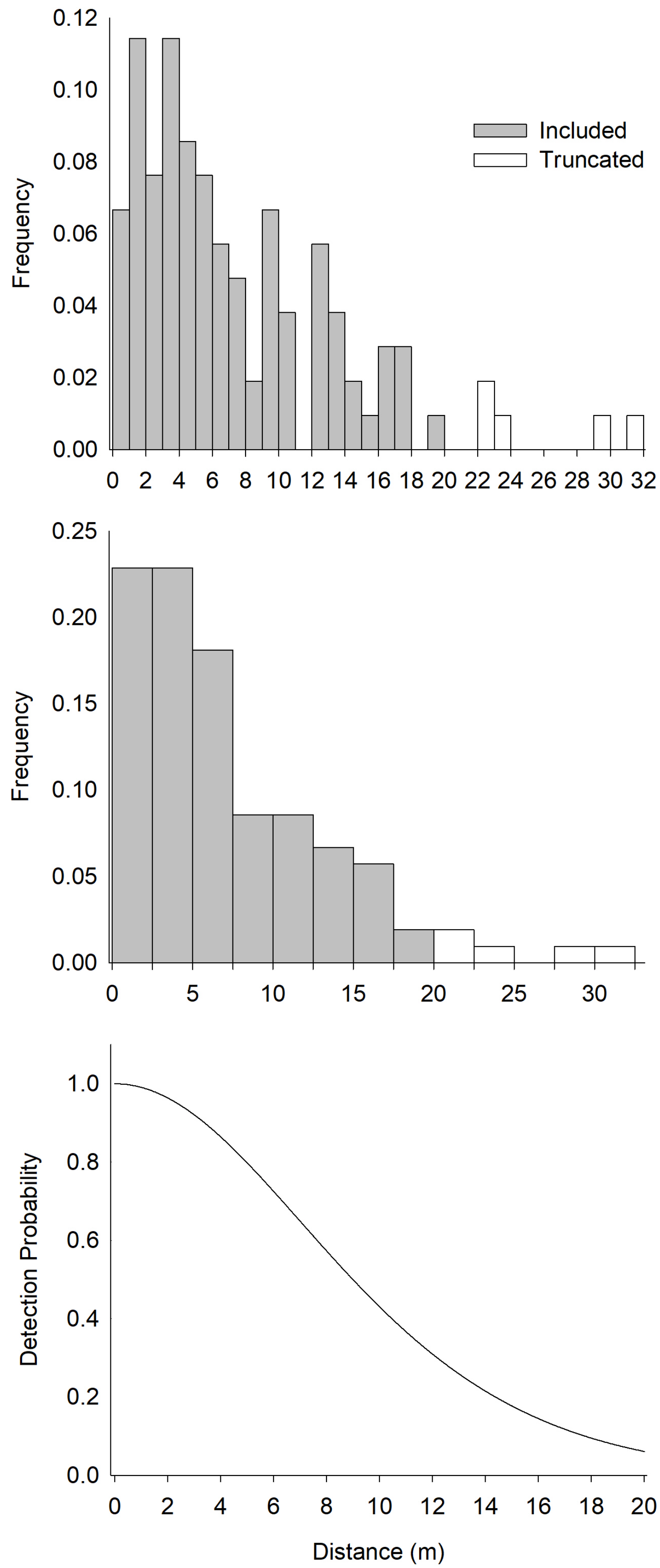
Fig. 4 Influence of four covariates on detection probability of the Pima Pineapple Cactus based on distance sampling along lines in southern Arizona, 2016-17. Estimates are from multiple-covariates distance sampling with half normal key functions and cosine adjustments. Estimates are shown at covariate levels equaled to the lower, middle, and upper quartiles. Inset box plots show distributions of each covariate

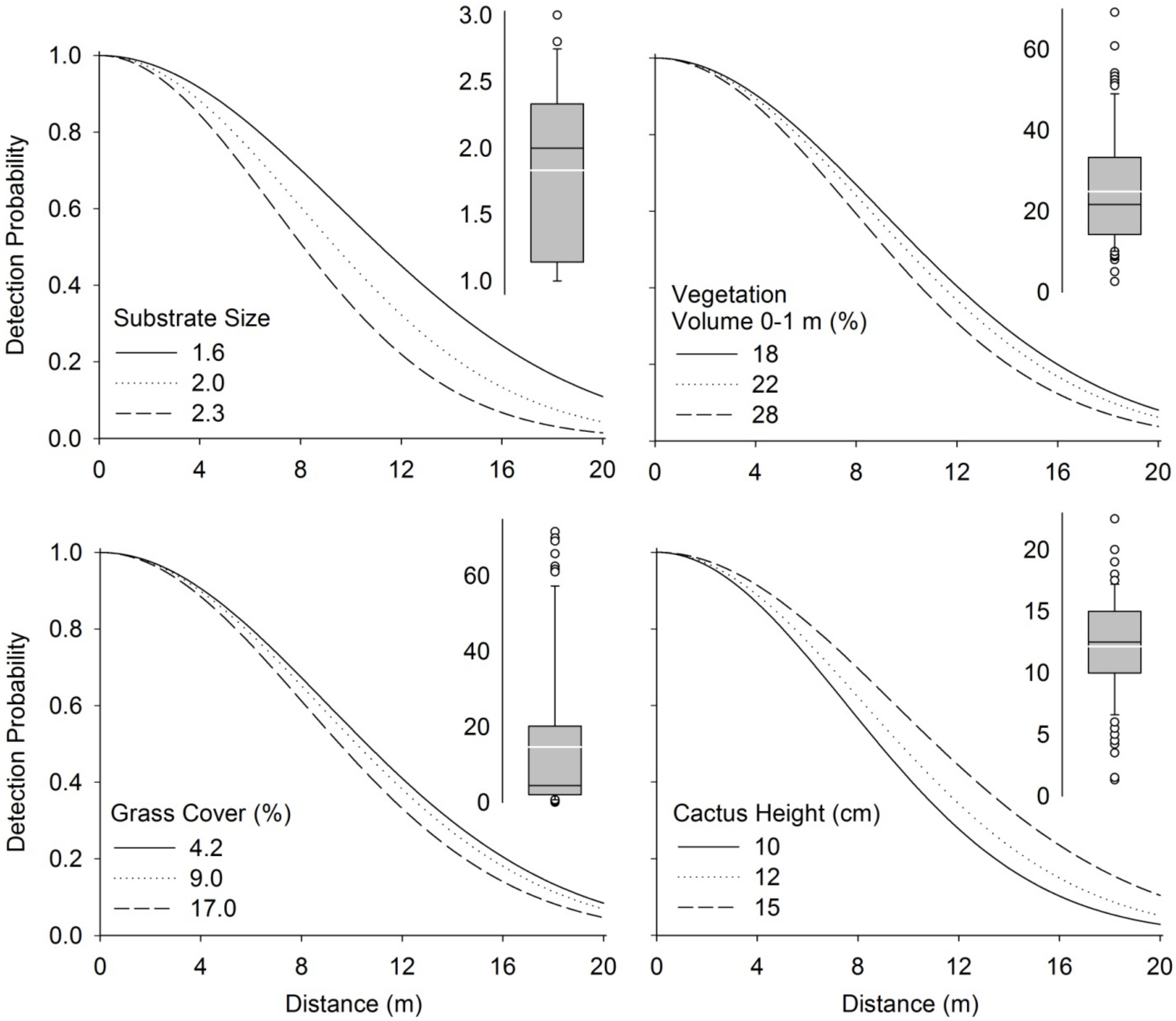
Fig. 5 Linear associations between estimated raw (no./ha; top) and log (bottom) densities of the Pima pineapple cactus in southern Arizona, 2016-17. Pearson correlation coefficients (r) are noted.

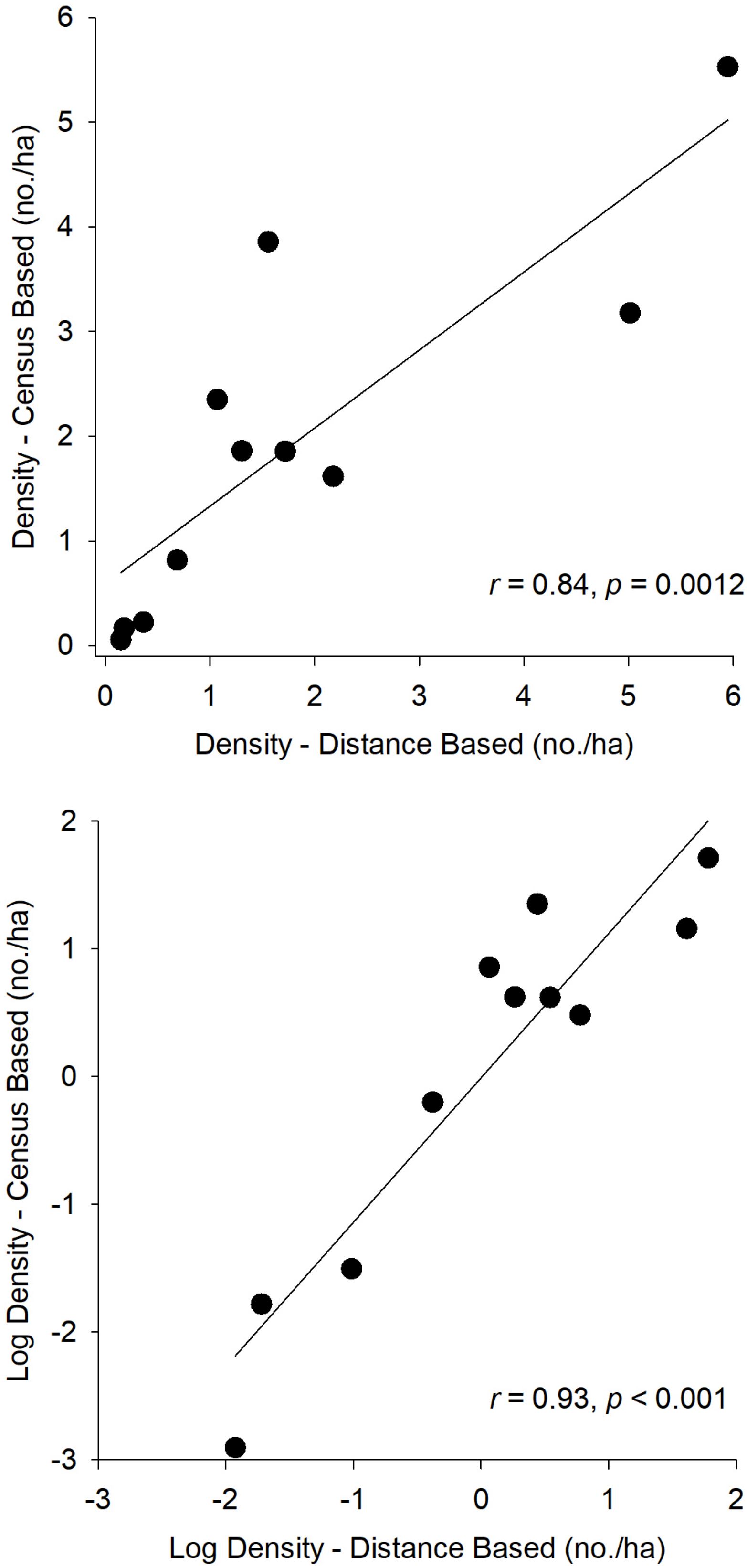

AZ
NM
Mexico
Tucson
Santa Cruz River
10
19
Cienega Valley
Altar Valley
Santa Cruz Valley
1
2
3
4
5
6
7
8
9
10
11
N
0
10 Kilometers
Mexico



2
1
6 m
Detected PPC
Undetected PPC
Environmental Sampling Points
Line Transect
Distance Plot Boundary
N
Meters
0
50
100


Frequency
Included
Truncated
Frequency
Detection Probability
Distance (m)


Detection Probability
Substrate Size
1.6
2.0
2.3
Vegetation
Volume 0-1 m (%)
18
22
28
Grass Cover (%)
4.2
9.0
17.0
Cactus Height (cm)
10
12
15
Distance (m)


6
5
4
3
2
1
0
Density - Census Based (no./ha)
r = 0.84, p = 0.0012
0
1
2
3
4
5
6
Density - Distance Based (no./ha)
2
1
0
-1
-2
-3
Log Density - Census Based (no./ha)
r = 0.93, p < 0.001
-3
-2
-1
0
1
2
Log Density - Distance Based (no./ha)

579