Rangeland Ecol Manage 65:330–339 | July 2012 | DOI: 10.2111/REM-D-11-00102.1

Forum

Using Very-Large-Scale Aerial Imagery for Rangeland Monitoring and Assessment: Some Statistical Considerations

Jason W. Karl,[1] Michael C. Duniway,[2] Sarah M. Nusser,[3] Jean D. Opsomer,[4] and Robert S. Unnasch[5]

Authors are [1]Ecologist and [2]Soil Scientist, USDA-ARS Jornada Experimental Range, Las Cruces, NM 88003, USA; [3]Professor, Department of Statistics, Center for Survey Statistics and Methodology, Iowa State University, Ames, IA 55011, USA; [4]Professor, Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA; and [5]Director of Science, Idaho Chapter, The Nature Conservancy, Boise, ID 83702 USA.

Abstract

The availability of very-large-scale aerial (VLSA) imagery (typically less than 1 cm ground-sampling-distance spatial resolution) and techniques for processing those data into ecosystem indicators has opened the door for routinely using VLSA imagery in rangeland monitoring and assessment. However, for VLSA imagery to provide defensible information for managers, it is crucial to understand the statistical implications of designing and implementing VLSA image studies, including consideration of image scale, sample design limitations, and the need for validation of estimates. A significant advantage of VLSA imaging is that the researcher can specify the scale (i.e., spatial resolution and extent) of the images. VLSA image programs should plan for scales that match monitoring questions, size of landscape elements to be measured, and spatial heterogeneity of the environment. Failure to plan for scale may result in images that are not optimal for answering management questions. Probability-based sampling guards against bias and ensures that inferences can be made to the desired study area. Often collected along flight transects, VLSA imagery lends itself well to certain probability-based sample designs, such as systematic sampling, not often used in field studies. With VLSA image programs, the sample unit can be an entire image or a portion of an image. It is critical to define the sampling unit and understand the relationship between measurements and estimates made from the imagery. Finally, it is important to statistically validate estimates produced from VLSA images at selected locations using quantitative data of the same scale and more precise and accurate than the VLSA image techniques. The extent to which VLSA imagery will be useful as a tool for understanding the status and trend of rangelands depends as much on the ability to build the imagery into robust programs as it does on the ability to quickly and relatively easily collect VLSA images over large landscapes.

Resumen

La disponibilidad de imágenes aéreas a gran escala (IAGE) (normalmente menos de un cm de de distancia de resolución espacial en el terreno) y técnicas que procesen esos datos dentro de indicadores del ecosistema han abierto la puerta para que de manera rutinaria se use IAGE en pastizales en monitoreo y evaluación. Sin embargo, para IAGE proveer información defendible para administradores es crucial para entender las implicaciones estadísticas para diseñar e implementar estudios de IAGE que incluyan consideraciones de escala de la imagen, limitaciones en el diseño de muestreo y la necesidad de validación de los estimadores. Una ventaja significativa de IAGE es que el investigador puede definir la escala (ejm. resolución espacial y extensión) de la imagen. Los programas de IAGE deberían planear escalas que empaten preguntas de monitoreo, el tamaño de los elementos del paisaje a ser medidos y la heterogeneidad espacial del medioambiente. Fallas en planear la escala puede resultar en imágenes que no son optimas en resolver las preguntas del administrador. Muestreos basados en probabilidad protegen contra sesgo y aseguran que la inferencia puede ser hecha para la area de estudio deseada. Seguido, recolección a lo largo de vuelos en transectos, IAGE permite bien a cierto diseño de muestra basado en probabilidad como diseño sistemático no usado a menudo en estudios de campo. Con programas IAGE la unidad de muestreo puede ser la imagen completa o una porción de ésta. Es fundamental definir la unidad de muestreo y entender la relación entre medidas y estimaciones hechas de la imagen. Finalmente, es importante validar estadísticamente los estimadores producidos de IAGE es lugares seleccionados usando datos cuantitativos de la misma escala y más precisos y certeros que las técnicas de IAGE. La amplitud a la cual IAGE será de utilidad como herramienta para entender el estatus y tendencia de los pastizales, depende en gran medida en la habilidad para construir imágenes en programas robustos sino también con la habilidad de recolectar imágenes IAGE rápidamente y relativamente fácil sobre grandes paisajes.

Key Words: accuracy, monitoring, precision, sample design, statistics, very-large-scale aerial image

INTRODUCTION

The importance of assessing and monitoring resource conditions and trends for rangeland management is well established (National Research Council 1994). As the uses of rangelands (e.g., grazing, recreation, energy development) become more diverse and threats to rangeland ecosystems (e.g., invasive

---

Presented at the SRM symposium, Very-High Resolution Imaging for Resources Monitoring, 9–10 February 2011, Billings, MT.

Correspondence: Jason Karl, USDA-ARS Jornada Experimental Range, New Mexico State University, Las Cruces, NM 88003, USA. Email: jkarl@nmsu.edu

Current Address: Michael C. Duniway, Research Ecologist, US Geological Survey, Southwest Biological Science Center, Canyonlands Research Station, Moab, UT 84532, USA.

Manuscript received 14 June 2011; manuscript accepted 19 March 2012.

species, wind and water erosion, climate change) become more pervasive, the need for robust monitoring and assessment programs that provide information at both local and national levels has become paramount. However, there are significant challenges to implementing monitoring or assessment programs that can provide useful information over very large landscapes.

Various remote sensing approaches have been proposed to accomplish consistent monitoring and assessment over large landscapes (Booth and Tueller 2003; Hunt et al. 2003). In particular, much attention has been paid to the development of complete maps of ecosystem indicators across large landscapes. However, most of these approaches are limited by the spectral and spatial resolution of commonly available sensors in their ability to provide the accurate and precise estimates of ecosystem attributes needed for long-term rangeland monitoring (Marsett et al. 2006).

Using very-large-scale aerial (VLSA) imagery—images taken with a ground-sampling distance (i.e., spatial resolution or on-the-ground dimensions of a single image pixel) of a few centimeters or less taken from aircraft—it is possible to generate reliable estimates of many attributes important for rangeland management decision making, such as vegetative cover (often by species; Booth et al. 2003, 2005a, 2005b; Booth and Cox 2008; Duniway et al. 2012), density, and canopy-gap size distributions (Karl et al. 2012). Estimates are derived from VLSA images using either manual image interpretation (e.g., Booth et al. 2006; Duniway et al. 2012) or automated classification techniques (e.g., Laliberte et al. 2006; Luscier et al. 2006; Karl et al. 2012). Given their small areal extent and large file sizes, VLSA images are typically not used to create maps of large landscapes. Instead, multiple VLSA images are acquired on a sample of land units from the overall study area.

The increasing availability of VLSA imagery and the development and validation of techniques for processing those data into indicators of ecosystem status and function has opened the door for VLSA imagery to be used routinely in rangeland monitoring and assessment. However, obtaining VLSA imagery for a large study area can be expensive compared to acquiring coarser-resolution imagery, and studies that rely on VLSA image–based indicators without proper measurement and sample design will be subject to various forms of error that can lead to biased or imprecise estimates of rangeland characteristics. Thus, to be used successfully in a monitoring and assessment context that provides defensible estimates for managers, indicators derived from VLSA imagery should be implemented within a statistically valid sampling design.

In this article, we discuss how to use VLSA imagery in the context of a sample survey of rangeland to provide a basis for characterizing a landscape at a point in time (i.e., assessment) or establishing a trend over time (i.e., monitoring). Accordingly, we consider the use of VLSA imagery from the perspectives of achieving unbiased estimates and minimizing sampling-related errors in estimates. We consider the following topics: 1) determining appropriate VLSA image scale; 2) designing VLSA-based sampling strategies for large landscapes; 3) identifying the sample unit and understanding the relationship between observations, measurements, and estimates with regard to VLSA images; 4) validating estimates from VLSA images; and 5) monitoring to detect change with VLSA imagery. We conclude with a set of recommendations for implementing VLSA imagery in rangeland monitoring and assessment efforts. This article touches on some important design considerations for monitoring and assessment with VLSA imagery but does not provide a comprehensive discussion of the topic.

A SAMPLING CONTEXT FOR VLSA IMAGERY

In remote sensing, as spatial resolution increases (i.e., pixel size or ground-sampling distance becomes smaller), the extent of the image typically becomes smaller. Most applications of remote imagery to rangeland monitoring and assessment have used either a single image or a small number of images mosaicked together to cover the entire study area, and information is extracted from the seamless image (Lillesand and Kiefer 1994). With VLSA image applications, however, the extent of each individual image is typically very small compared to the entire study area, and it often requires extensive manual processing or custom workflows to produce accurate mosaics of large areas from VLSA images (see Laliberte et al. 2010). Instead, individual VLSA images are often acquired only at selected locations within a study area. Estimates for the larger area are then made from this sample of images.

While some of the considerations below will apply to estimates derived from mosaics of many VLSA images (sensu Laliberte et al. 2010) or remote sensing in general, our primary focus is the application of VLSA images to rangeland monitoring and assessment in this survey sampling context. We use the term *survey sampling* (or simply *sampling*) to mean randomly selecting a sample based on a probability sample design. In such a design, the probability of selecting a land unit or image is specified by the design prior to making the selection. Defining the selection probability via a random sampling design provides the basis for scientifically valid estimates of bias and precision. We specifically recommend against haphazard, purposive, or convenience samples whose bias and precision cannot be credibly quantified.

Below we discuss factors that affect measurements or estimates in terms of accuracy, precision, and cost/time efficiency of sampling with VLSA imagery. As will be described further below, when images or land units are selected through probability-based sampling, a rich set of statistical tools is available to produce estimates and assess their accuracy and precision without the need to develop models for the observed variables. In this context, *accuracy* refers to how close estimates or measurements are to the actual value. The concept of statistical accuracy is composed of two related but distinct components: bias and precision. The *bias* of an estimator is the difference between the average of the estimates over repeated sampling and the actual value in the field. *Precision* refers to how close repeated measurements or estimates are to each other regardless of the actual value in the field. The precision of a measurement is estimated by the *sample variance*, which is a measure of variation between sample measurements. Under simple random sampling, the sample variance is denoted as s^2, and the *sample standard deviation* is the square root of the sample variance, or s (the formula for s^2 is different when other probability designs are used). The precision of an estimate is estimated with the *standard error*. Under simple random sampling, the standard error is $s/n^{1/2}$. In statistical practice, an estimate is considered good if it has no or negligible bias and a level of precision that can be suitably quantified by one of these measures. We refer to Lohr (2009, chap. 2.2) for a more detailed description of these concepts.

DETERMINING APPROPRIATE SCALE

One significant aspect of using VLSA imagery for rangeland monitoring and assessment is selecting the appropriate scale of imagery to provide the best estimates as cost efficiently as possible to answer management questions. Scale has many meanings and implications in ecological sciences (Addicott et al. 1987; Turner et al. 1989; Kotlair and Wiens 1990; Farina 1998), but most relevant here are the concepts of extent and grain of the images. *Extent* refers to the total ground area captured by a VLSA image as defined by the maximum dimensions of a single image. *Grain* refers to the size of the smallest unit of observation within an image (i.e., pixel ground sampling distance). The choice of the extent and grain is important for using VLSA images for rangeland monitoring and assessment because it affects what can be reliably detected from the image and the variance of observations observed from VLSA images (Woodcock and Strahler 1987).

A significant advantage of VLSA imaging over other sources of remote imagery (e.g., LANDSAT or NAIP imagery) is that the researcher can specify the extent and grain of the resulting images. With VLSA imagery, altitude of the aircraft and zoom level of the camera can be specified to achieve specific extent and grain sizes. Collection of VLSA imagery from light aircraft (Booth et al. 2003) or unmanned aerial vehicle (Rango et al. 2009) can offer even more flexibility in altitude to customize image properties because they can safely fly lower than regular piloted aircraft. Likewise, for nadir-looking photos taken in the field using a camera stand (e.g., Booth et al. 2004), zoom level and height of the camera can also be controlled.

Some a priori knowledge of the landscape being assessed or monitored is necessary to determine the best image resolution and extent combination for a particular monitoring objective. The scale of the monitoring question, size of landscape elements to be measured or observed, and spatial heterogeneity of the environment should drive the selection of resolution and extent of VLSA images.

Selecting Spatial Resolution

Because of the influence it has on the accuracy and precision of rangeland attribute estimates, selecting an appropriate spatial resolution for VLSA imagery is important. Strahler et al. (1986) described two ways of representing landscape objects in an image that are based on the relationship between the spatial resolution of the image and the size of the objects. A low-resolution view occurs when the individual image pixels are larger than the objects of interest (Fig. 1). This can result in many pixels that overlap different objects (i.e., "mixed" pixels) and a low ability to resolve individual objects or cover types. Alternatively, a high-resolution view occurs when the elements in a scene are larger than the individual image pixels. In this case, it becomes possible to detect edges and spatial arrangements of objects because the likelihood of pixels overlapping different objects is reduced. In the case of VLSA imagery, each pixel represents such a small amount of area on the ground that there may be almost no mixed pixels. While this can increase the ability to resolve landscape elements and cover types, excessive resolution can also be problematic.

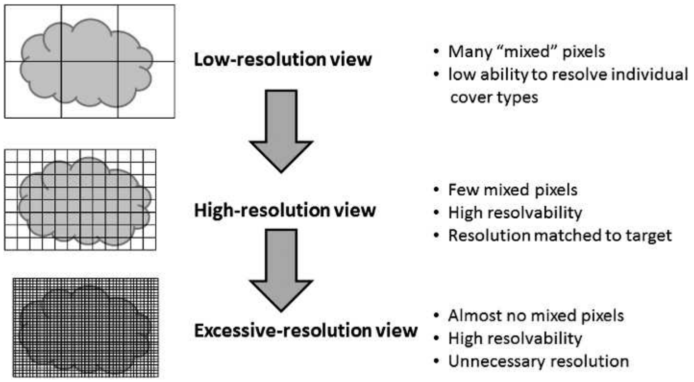

Figure 1. The relationship between the size of an object on the ground and the resolution of an image of it determines whether the image constitutes a low- or high-resolution view of the object (Strahler et al. 1986). Matching image resolution to the features being measured ensures the best results while avoiding negative aspects of excessively high resolution like high fine-scale heterogeneity (e.g., noise) or large computer storage and processing requirements.

A higher spatial resolution than necessary can make it more challenging to work with and store VLSA imagery and may not increase the ability to resolve landscape objects or cover types over a coarser resolution. The large file sizes typical of VLSA images can be difficult to store and process on current desktop computers. Spatial resolution higher than necessary to detect patterns or objects of interest also poses problems for automated classification of imagery because of increased heterogeneity within the image, as variance within important landscape elements is represented by variance among pixels. However, this is less of a problem with approaches that use interpretation of multiple points on an image (sensu Booth et al. 2006; Duniway et al. 2012). Finally, for a given camera lens or sensor, spatial resolution is inversely related to image extent. Higher spatial resolution than necessary to detect the landscape objects of interest can result in image extents that are too small to effectively sample across environmental heterogeneity (see Selecting Extent below). For these reasons, it is worthwhile to carefully consider (and even experiment with) what are appropriate spatial resolutions for a given monitoring objective.

Selecting Extent

Image extent is determined by the altitude of the sensor, the focal length of the sensor lens, and the dimensions of the sensor itself (Lillesand and Kiefer 1994). Image extent should be determined by the spatial heterogeneity of the landscape and the monitoring objectives. In patchy environments, sampling units (e.g., plots or subplots) that are smaller than the patch size can lead to many plots with extreme values and high standard errors when averaging over many units (Elzinga et al. 1998). Increasing the size of the sample unit so that fine-scale

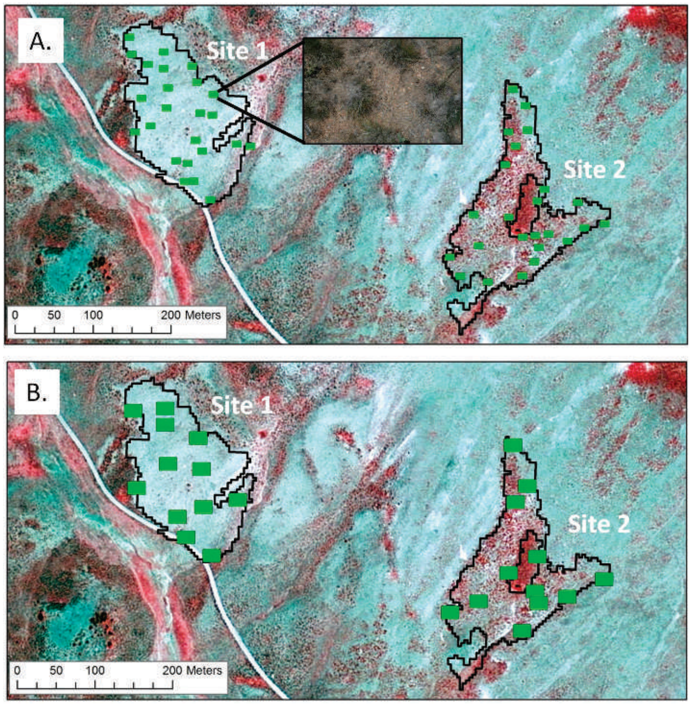

Figure 2. Extent of VLSA images should be adjusted to sample across heterogeneity within a site. Estimates of sample size needed to detect a 50% change in percent shrub cover were higher from 24 downward-looking photos taken from 2-m above the ground (**A**, inset photo provides an example) and were higher for site 2 because shrub patch patterns were larger than the image extents than for site 1, which was homogeneous grassland. By increasing the extent of the images (**B**), however, sample size estimates were greatly reduced for site 2. See Table 1 for details.

environmental heterogeneity is captured within a sample unit rather than between sample units can reduce standard errors of estimates.

For example, the two areas shown in Figure 2A were each sampled with nadir-looking photographs taken at 2 m above ground level, yielding an approximate extent of 1.4×0.9 m and ground-sampling distance (e.g., spatial resolution) of about 0.5 mm. Twenty-four photographs were taken in each area. Note that the background color infrared aerial photograph suggests that these two areas differ in the amount of shrub cover and that the spatial heterogeneity of the sites is different—site 1 is very homogeneous, whereas site 2 shows shrub patches of varying densities. These properties have implications for sampling the sites and the ability to detect differences between them. Using the program SamplePoint (Booth et al. 2006), shrub cover was estimated for each image. The standard error for percent shrub cover in site 1 was lower than for site 2 (Table 1): 1.67% and 6.55%, respectively. From the image-derived cover estimates, the estimated number of samples needed to detect a 50% change in shrub cover in site 1 was 24, whereas for site 2 it was 66.

Larger extent images for these two sites were simulated by averaging together shrub cover estimates from pairs of closest images so that there were 12 images per site (Fig. 2B). With the larger extent images, the standard error for site 1 was 1.87% and for site 2 was 4.86%. Estimated number of samples to detect a 50% change in shrub cover decreased to 16 for site 1 and 19 for site 2. Site 1 was relatively homogeneous with respect to shrub cover, resulting in a relatively small increase in standard error going from the small to larger extent images. Conversely, site 2 had a much patchier shrub distribution, and thus larger extent images had a much lower standard error than the small images, and subsequently fewer samples were needed to detect change in that area. In both areas, however, using a larger extent that sampled over the fine-scale heterogeneity of the sites resulted in shrub cover estimates with lower standard errors.

IDENTIFYING THE SAMPLE UNIT

When using VLSA imagery for monitoring and assessment, careful consideration must be given to defining the statistical population, clearly identifying what constitutes a sample unit, and how inferences to the larger population are to be derived from measurements within the sample units. A (statistical) population is the entire set of entities to which conclusions (i.e., inferences) are to be drawn from the sample data that are collected. For rangeland studies, the population is often defined by the extent of the landscape of interest. Any successful monitoring or assessment program must have the statistical population explicitly defined because it determines the bounds for sample selection and for defining the sample units that can be selected for measuring. A sample unit is an object or area that will be selected for measurement or observation. In rangeland studies, the sample unit is typically a plot—a relatively small homogeneous region of land in the study area that is characterized by one or more measurements.

An observation is a single measurement associated with the sample unit. In some cases, there may be a single measurement per sample unit, such as counting the number of juniper in an aerial photograph or classifying juniper from an image, as in Strand et al. (2006) or Laliberte et al. (2004). However, it is often not possible to make measurements for the entire sample unit because 1) the sample unit is too large to sample effectively for a particular attribute with a single measurement (e.g., density of a rare species) or because 2) the "true" value of the attribute cannot be known or precisely quantified (e.g., cover of perennial grasses) at the scale of the sampling unit. In both of these cases, multiple measurements are typically taken by subsampling within the sample unit and combined (e.g., through simple averaging, summation, or other techniques) to provide an estimate for the sample unit. Subsampling may involve random location of the subsampling observation unit (e.g., a transect) or consistent location of the observation unit within the plot (e.g., transects are centered on the centroid of the plot and oriented consistently). For example, the measure of interest may be canopy height, and the observation unit within the sample plot may be a randomly located transect along which 50 observations of canopy height are taken at systematic intervals. These observations can be averaged to provide an estimate for the transect that represent an estimate of canopy height for the plot. In some studies, the combined value is assigned to the sample unit, while in others the within-sampling-unit variance is also recorded.

When using VLSA imagery as a source for monitoring data, the relationship between sample unit and observations can be

Table 1. Comparison of percent shrub cover estimates from two different sites where the estimates were derived from images of different extent (see Fig. 2). Estimates for the small-extent images were made from 24 nadir-looking photos taken at 2 m above ground level. The larger-extent image values were simulated by averaging pairs of adjacent small-extent images.

	Sample average	Sample standard deviation	Sample standard error	Coefficient of variation	Estimated sample size to detect 50% change[1]
Sample of 24 small-extent images					
Site 1	13.49%	8.18%	1.67%	0.606	24
Site 2	31.39%	32.08%	6.55%	1.022	66
Sample of 12 larger-extent images					
Site 1	13.49%	6.46%	1.87%	0.479	16
Site 2	31.39%	16.84%	4.86%	0.536	19

[1] Sample size requirements (*n*) for a two-sample *t* test estimated as $n=2S^2(Z_{\alpha/2}+Z_{\beta})/\mathrm{MDD}^2$, where *s* is the standard deviation of the sample, *Z* is the standard normal coefficient using $\beta=(1-$ probability of a type II error$)=0.8$, $\alpha=($probability of a type I error$)=0.1$, and MDD was the minimum detectable difference set to 50% of the current mean (Elzinga et al. 1998; Herrick et al. 2009). No finite population correction was applied.

different, depending on the extent of the images related to the plot and how the measurements or observations will be made. It is recommended that the extent of the image be larger than that the plot (e.g., Booth et al. 2003; Duniway et al. 2012). In this case, measurements for the sample unit can be made directly from the image, or various interpretation (e.g., Booth et al. 2006; Karl et al. 2012) or classification (e.g., Laliberte et al. 2006; Luscier et al. 2006) approaches can be used to derive an estimate for the sample unit. If the extent of the images is much smaller than the plot, then measurements or estimates made for each photo must be expanded to generate an estimate for the whole plot (i.e., sample unit). In this case, care must be taken to ensure that enough measurements are taken (i.e., enough VLSA images and/or enough measurements per image) to derive reliable estimates for the plot and that subsamples are appropriately located within the plot (see above). An advantage of many VLSA-based methods over other remote sensing techniques for measuring rangeland attributes is that much more control over image properties is available so that users are able to match the scale of imagery to their specific needs.

VLSA-BASED SAMPLE DESIGN FOR LARGE LANDSCAPES

The aim of sampling for monitoring and assessment is to be able to draw conclusions about the condition or trend of a larger area through measurement or observation of a small subset of sites (Lohr 2009). In this section, we focus on sample design considerations for aerial image applications. Field-based image collection (e.g., downward-looking photographs taken from a camera stand; see Booth et al. 2004) is similar to other field-based collection techniques with respect to sample design, and guidance exists for these situations (e.g., Elzinga et al. 1998). VLSA image sampling, however, poses different constraints and opportunities for sampling.

Probability-based sampling helps guard against site selection bias and supports making generalizations to a larger, defined landscape. Two properties of probability-based sampling ensure that the scientific validity of the estimates do not rely on subjective decisions by researchers, which would make them vulnerable to criticism. First, if locations for measurement are selected using a probability sample from the entire target study area, statistical estimation methods provide unbiased estimates of indicators (e.g., percent bare ground) that are representative of the whole study area or regions within the study area. Second, valid measures of the precision of the estimates can be calculated.

But is probability-based sampling economically and logistically feasible with VLSA image based monitoring and assessment? Practical considerations, such as targeting sample locations while maintaining geographic spread, choosing appropriate sampling units, obtaining access to sampled locations, and minimizing travel costs, present challenges when designing such studies.

Modern probability sampling methodologies provide a natural framework for balancing statistical and logistical considerations of sample design with study objectives (Cochran 1977; Thompson 2012; Gregoire and Valentine 2008; Bethlehem 2009; Lohr 2009). A variety of sampling units may be used in selecting a sample, including land areas or points on the land. For the kinds of studies we are concerned with, multitiered sample designs (e.g., stratified sampling, two-stage or multistage sampling, cluster sampling) are often needed to address scientific objectives and practical constraints.

Sampling Considerations

Ideally, a probability sampling design is applied to the full target population as a foundation for unbiased assessments of characteristics of interest for the entire area as well as for domains within that area, where a domain is a subset of the population that can be defined along a measured characteristic (e.g., all grasslands within the study area). A variety of methods may be used to randomly select the sample locations, including simple random sampling, systematic sampling, stratified random sampling, unequal probability sampling, and cluster sampling.

Under simple random sampling, a set of sample units is randomly selected without restrictions placed on the randomization process, and each unit has an equal chance of being selected. Simple random sampling has been shown to perform adequately for maps of many terrains. However, because of the lack of control in the selection process, the sample size for individual land cover categories may be too small to support useful inferences, and the selected locations might be spread in a way that is expensive to collect data with VLSA image–based approaches.

Systematic sampling involves selecting a random start and then using a fixed interval to select sample units. It generally provides better geographic spread than a simple random sample but does not ensure adequate sample sizes for subpopulations. Systematic sampling can be challenging to implement in a field-based survey of a large landscape because of difficulties and cost (i.e., time) in getting to the sample locations. With a VLSA image survey, however, systematic sampling may be the easiest approach to sampling a large landscape. Once a random start location has been selected, the other sample locations can be arranged along parallel flight lines to efficiently collect all the imagery required. With a systematic VLSA image survey, there is no need to randomize orientation of transects if the start location of one transect is randomly selected and images are acquired at fixed intervals along all transects. This permits orientation of transect flight lines to maximize image quality or efficiency of collection.

A third sampling procedure is selection with probability proportional to a size or importance measure, called pps sampling. Unlike systematic and simple random sampling, sampling units in pps are assigned selection probabilities that are proportional to an importance factor (e.g., size of sampling unit, levels of interest in specific domains) that results in "more important" units having a higher selection probability. Although pps sampling can be advantageous in targeting the sample composition, it does not guarantee adequate sample sizes for subpopulations. On the other hand, it can greatly increase the efficiency of field-based sampling because effort is focused on those areas that are considered "more important." A pps sampling approach could also be relatively efficient for VLSA image surveys if the sample locations are concentrated in their distribution.

Stratified random sampling is often used to allocate the sample across subpopulations in a more controlled fashion. The target population (i.e., study area in this case) is divided into mutually exclusive areas (i.e., strata), corresponding to subpopulations, such as land potential units (e.g., ecological sites), land cover categories, or simply geographically defined regions regardless of land cover. Independent samples are selected from each stratum using a probability sampling procedure (e.g., simple random, systematic, pps). It is not necessary to use the same design for each stratum, providing flexibility in addressing special operational constraints or subject matter objectives. For example, areas that are expected to experience a high degree of change may be placed in a stratum and a separate design developed to improve ability to detect small amounts of change (e.g., by increasing sampling intensity). Appropriate strata for rangeland monitoring and assessment are often related to vegetation communities (e.g., ecological sites) or management units (e.g., grazing allotments), but for temporal monitoring, it is often best to use geographically defined strata. Regardless, stratification typically increases the precision of estimates and may lead to cost efficiency gains.

For VLSA image surveys, it is often most efficient to collect images along long flight lines, which can restrict the types of stratification that can be used. In this case, an alternate approach for VLSA image surveys could be to employ a systematic sampling strategy along flight lines with sufficient sampling intensity to support poststratification of the sample data (Sarndal et al. 2003; Lohr 2009). Poststratification consists of adjusting the sampled counts for a given set of categories (the "poststrata") so that they match known population counts or areas. Poststrata for rangeland studies might be associated with regions where the total or proportion of surface area for each region is used in the adjustment. This approach is frequently used in survey estimation, and it is known to improve the precision of estimators by grouping units into more homogeneous poststrata, similarly to what would happen if the sample were drawn according to a stratified design with the same set of categories.

Another sampling strategy that could work well with VLSA imagery is cluster sampling. In cluster sampling, the original sample units (plots) are combined into clusters, or collections of nearby observations. From a statistical perspective, clusters are ideally defined as a set of plots that mimic the entire population (or stratum), but this is rarely feasible. Cluster sampling is typically used to improve operational efficiencies. For example, flight costs per sample unit could be reduced by selecting clusters of adjacent sampling units. Cluster sampling can be used in concert with stratification by stratifying clusters into homogeneous groups, and individual clusters may be selected via simple random sampling, systematic sampling, or pps sampling. The disadvantage of cluster sampling is that in practice, the data for adjacent observations are often more similar to each other than for observations that are further apart. Because individual units within clusters are correlated, for the same total overall sample size, a sample of clusters contains less information about the target population than when the data come from a simple random sample of individual units. This loss of precision can be offset by increasing the number of selected clusters, which is feasible if the cost savings generated by the clustering are sufficiently high. Practitioners often calculate anticipated variances of estimates for alternative designs prior to settling on a final sample design for field data collection.

VALIDATION OF VLSA ESTIMATES

Ensuring that estimates derived from VLSA imagery accurately reflect the attribute being considered is a crucial step in using VLSA imagery in rangeland monitoring and assessment programs. This process, called validation, is often overlooked or only minimally treated in monitoring programs. Failure to properly validate estimates of rangeland attributes can at best hamper the ability to detect change and at worst lead to erroneous conclusions. Validation is more than just an accuracy assessment of VLSA estimates. It is a process of determining whether VLSA estimates are suitable for a given objective, which includes determining whether the accuracy and precision of VLSA estimates is suitable for monitoring purposes and, if not, what can be done about it (e.g., improving the estimates through modified estimation techniques or additional training data or developing adjustments to correct for over- or underestimation; see below). Validation provides information necessary to judge the utility of VLSA estimates for use in rangeland monitoring and assessment programs and the reliability of a given technique for deriving VLSA estimates.

What Is an Appropriate Validation Data Set?

Several criteria should be considered when selecting a data set for validation of rangeland parameters estimated using VLSA. The validation data must be independent of the VLSA estimates. For example, collecting the data a second time using the same method on the same image would not be appropriate for validation (this would test the repeatability of the method). The accuracy and precision of the validation data must be substantially greater than the VLSA-derived estimates. For example, ocular estimates of vegetation cover have been shown to have low precision and accuracy and thus would not be appropriate for validation (Godinez-Alvarez et al. 2009). Ground data collected with standard quantitative rangeland methods are often the best choice for validation. The validation data should be collected in a manner that is consistent with the method used to collect rangeland parameters from the VLSA imagery. For example, estimates of cover obtained using object-oriented image analysis software packages or point image interpretation techniques conducted on coarser resolution VLSA measurements ($\sim$3–5-cm pixel) are often more akin to canopy cover than foliar cover measured using common point intercept methods (e.g., Karl et al. 2012). When methods of the VLSA estimate and validation data set differ, it is likely that there will not be a 1:1 correlation (e.g., Duniway et al. 2012). Finally, the ideal validation data set would be collected using methods commonly employed by many land management and resource monitoring agencies, such as the USDA-NRCS National Resource Inventory program. Selecting such methods will allow estimates of rangeland parameters from VLSA to be integrated with existing data sets for analysis.

The scale of the validation data should also match the scale of the VLSA estimate both in resolution and in extent (Fig. 3). In this context, resolution refers to the spacing of the individual measurements or observations taken within the image or plot area. Validating VLSA image–based estimates with field data that are of much lower resolution (Fig. 3A) can potentially result in biased and low-precision estimates for the validation data set. Field data that are of a different extent can yield validation data estimates that do not match the area being estimated or measured from the VLSA image (Figs. 3B and 3C). Validation data sets that have similar extents and resolutions as the VLSA image–based estimates (Figs. 3D and 3E) will provide the best validation of VLSA-based estimates because the area being sampled and the estimation techniques are as similar as possible.

Depending on the monitoring objectives and what aspect of the VLSA image estimation procedure is being validated, coregistration (i.e., matching of geographic locations between two data sets) of field data with the VLSA imagery may or may not be necessary. Exact coregistration of field validation data with VLSA imagery is exceedingly difficult because of the small size of VLSA image pixels (Weber et al. 2008) and the lack of precision of most GPS units (i.e., positional error of coordinates obtained from high-precision GPS units can still be many times larger than pixel size of a VLSA image). If the goal of a validation exercise is to evaluate the ability of an observer to identify a single, specific cover type using a point-based estimation technique, then precise coregistration of the field and image data may be needed. However, in most cases, it is sufficient to compare plot-level (or transect- or frame-level) estimates with field-based estimates or measurements of the same area. In this case, precise coregistration of the VLSA image with the field data is not necessary because the sample unit is the plot (or transect or frame) and not any individual observation or measurement within that area. This is true, however, only as long as the plot boundaries as determined in the field can be reliably identified on the VLSA image.

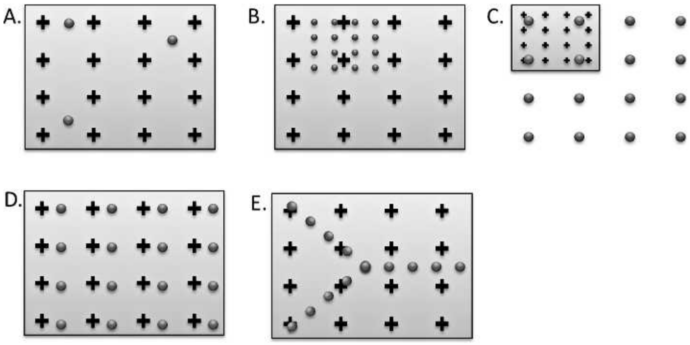

Figure 3. Validation of VLSA image–based estimates with field observations must be performed at the same scale. Differences in resolution (**A**), extent (**B**), or both (**C**) can lead to discrepancies between estimates of the same parameter due solely to scale. Matching scale, however, does not mean that field and VLSA image measurements must be taken from exactly the same locations, only from the same area at the same scale (**D, E**).

Determining Validity of VLSA Estimates

VLSA rangeland parameter estimates and validation data parameter estimates should be compared using a statistical approach that provides tests of both the accuracy and the precision of the VLSA estimates (Fig. 4). The goal of validation is to understand the overall accuracy of the VLSA-based estimates and understand if and when rangeland attributes are being under- or overestimated. To do this, it is not necessary to perform validation on every VLSA image. Rather, validation should occur on a subset of images that are representative of the range of ecological sites and conditions of the larger monitoring program (Duniway et al. 2012) or on a random sample of images if the study is large enough.

Regression approaches are ideally suited for comparisons of continuous data, such as percent cover and composition (as in Duniway et al. 2012). With the validation data used as the independent variable and the VLSA estimate as the dependent variable, both the root mean square error of the model fit and the confidence interval around the predicted regression line provide an estimate of the precision of the VLSA estimate in the units of the measurement. Confidence intervals around the estimated slope and intercept of the regression equation provide a means for evaluating the accuracy of the VLSA estimates. If the confidence interval of the intercept parameter estimate does not include 0 or the confidence interval of the slope does not include 1, then the VLSA estimates differ from the validation data in a systematic or biased manner (relationship with the validation data is not 1:1). For instance, when the VLSA estimates appear to be sufficiently accurate but not precise, increasing the samples size of VLSA estimates can usually overcome this lack of precision. In instances where the

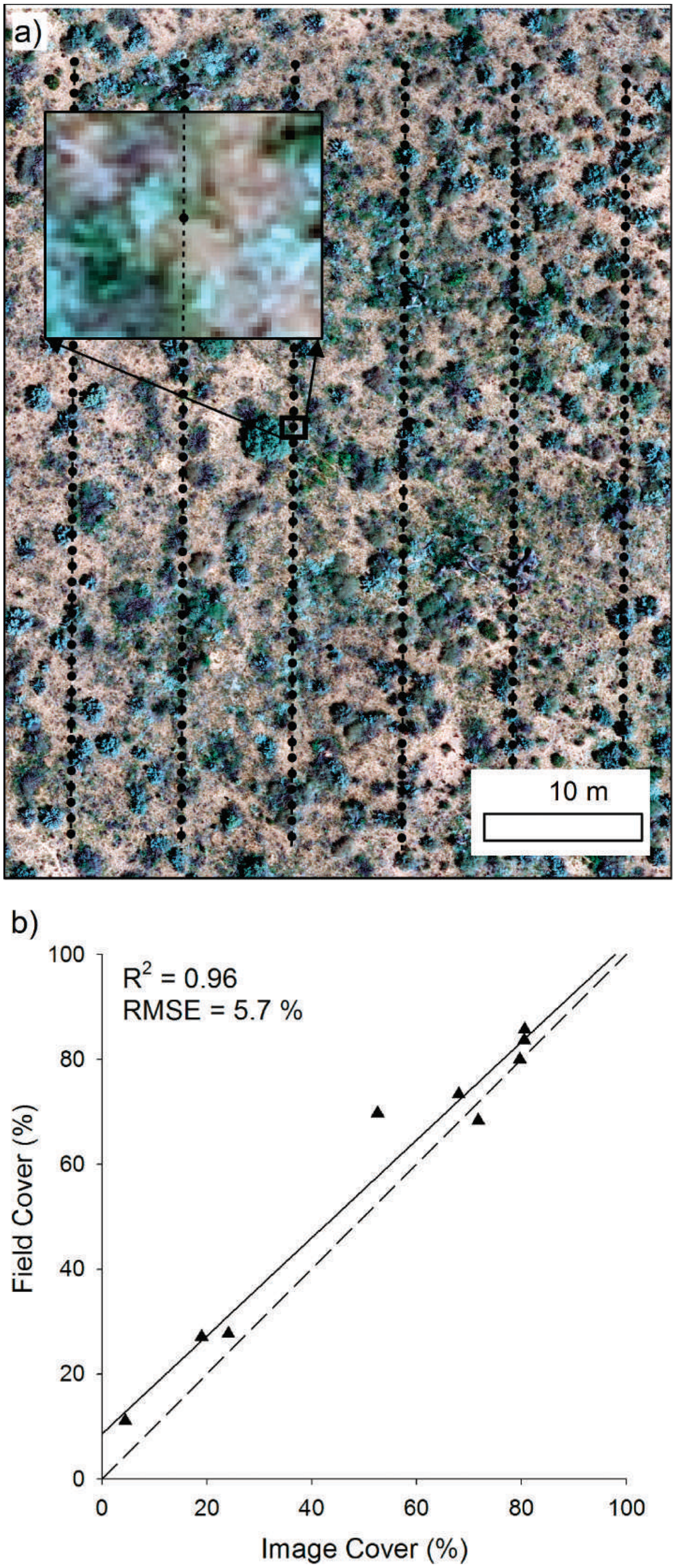

Figure 4. Regression between field and VLSA image estimates is one technique for validating VLSA image–based estimates.

VLSA estimates are precise but not sufficiently accurate, VLSA estimates can be adjusted to match the validation data using the regression parameters (Fig. 4). A significant downside to this solution is that the validation data are no longer independent of the VLSA estimates, and an additional validation data set will be needed to validate the new, adjusted VLSA estimates.

MONITORING TO DETECT CHANGE WITH VLSA IMAGERY

Monitoring and assessment, two terms that are frequently used interchangeably, are actually two distinct approaches to evaluating resource condition and trend, and appreciating the difference is important to successfully using VLSA imagery for monitoring purposes. Assessment is the estimation of the value of an ecological attribute or the functional status of an ecological process (Herrick et al. 2009). Assessments can employ either qualitative or quantitative methods but generally consist of quick-to-implement methods and emphasize sampling of a large number of sites at the cost of some precision at any single site. Assessments are frequently done to support management planning or decision-making efforts. Monitoring, on the other hand, is the orderly and quantitative collection and interpretation of resource data to evaluate progress toward a management objective (Herrick et al. 2009). Monitoring must be conducted over time, and because the goal is to detect changes in condition, precision requirements are higher for estimates at any given site. Successful monitoring programs require a high level of attention to specifying objectives, sample design, minimizing error, and achieving high-precision estimates (Elzinga et al. 1998), more so than assessment campaigns. Because of the greater emphasis on design and precision, it follows that monitoring is more than just sequential assessments.

Within a rigorously defined monitoring program, VLSA image–based techniques may offer some distinct advantages over purely field-based monitoring but will also have some limitations. An example where VLSA image techniques could benefit a monitoring program is achieving adequate sample sizes to detect the desired amount of change. Field studies are expensive, and it can be difficult to sample enough sites to have the statistical power to detect the desired amount of change. This problem is compounded in large, heterogeneous landscapes. With VLSA-image acquisition, it may be feasible to sample a much larger number of sites and improve the power to detect change.

One potential challenge to implementing VLSA image approaches for monitoring might be achieving the desired precision of estimates and keeping that precision consistent across different ecological sites. Duniway et al. (2012) reported that both the precision and the accuracy of cover estimates for perennial grasses and forbs were higher in some vegetation communities (e.g., mesquite communities in southern New Mexico) than for others (e.g., sagebrush steppe communities in Idaho and Nevada) because of differences in the amount of litter and other vegetation (e.g., annual grasses) that made it difficult to reliably discriminate between some plant functional groups. Whereas precision of point-based cover estimates from image interpretation can generally be improved by increasing the number of points, this will not help if there is confusion between similar-appearing cover types.

When one of the goals of a survey is to detect change over time, it is often beneficial to coordinate the sampling designs across time periods. If the monitoring goal is only estimating change in a target population or attribute, the most efficient estimates of that change are obtained if the sampling units remain the same across the time periods. In most surveys, however, estimating change is not the only goal, and in that case, efficient sampling designs will often be composed of a set of sampling units that overlap across time periods, while other sampling units will vary over time. The National Resources Inventory conducted by the USDA-NRCS is an example of a natural resource survey that is set up to estimate conditions at a point in time and how they change over time (Nusser and Goebel 1997). Its sampling design is composed of a core panel of plots that are revisited each year, supplemented by additional panels that are revisited at longer intervals (Breidt and Fuller 1999; Fuller 1999). The primary panel ensures that year-to-year changes can be estimated at a high level of precision for the target variables of highest interest. The supplementary panels also provide some information on change over longer time periods but are used mainly to give estimates of land characteristics of interest for individual years.

While keeping sample units fixed over time in surveys designed to estimate change is ideal from a statistical standpoint, there are major challenges to such an approach with VLSA-based sampling. With VLSA image acquisition, it may be challenging to acquire images with exactly the same properties (e.g., resolution, illumination) in different time periods. For estimating change, it is necessary to sample the same sample units over time (Nusser et al. 1998). While the use of GPS can help ensure that images are acquired from roughly the same spot, duplicating the exact properties of an image that affect the scale and quality of the image (e.g., altitude, orientation, off-nadir angle) from one time period to another is difficult. Given that differences in scale can affect the accuracy and precision of VLSA image–based estimates (see above), differences in images between time periods could lead to slight variation in the quality of estimates for the sample units. Another potential problem, not restricted to VLSA image approaches, is that the underlying population of interest might change, so that while the sampling units remain the same, they no longer track features of interest. For instance, a survey might target wetlands, and a site might have changed to deeper water or dried up, so that it is no longer part of the population. If the sampling units are periodically changed, then these less interesting units are removed from the sample and replaced by new ones.

MANAGEMENT IMPLICATIONS

We make the following recommendations when implementing VLSA imagery for rangeland monitoring and assessment. First, plan for an image scale (i.e., spatial resolution and extent) that matches monitoring objectives. Keep in mind that all VLSA images have an inherent scale, and if you do not specify your scale requirements, you may be left with images that are not optimal for answering management questions. Second, identify the sampling units relative to the VLSA imagery being used and understand the relationship between observations, measurements, and estimates of attributes within the sampling units. Third, avoid selection bias in selecting locations for acquiring VLSA imagery by relying on probability-based sample design techniques. Fourth, validate your VLSA-based estimates using data that are more precise and accurate than your VLSA-based techniques. Also, be sure to validate at the same scale as your estimates. Ideally, use standard field methods for validation so that data can be incorporated with other monitoring and assessment efforts (e.g., USDA-NRCS 2009; Mackinnon et al. 2011). Finally, consider long-term requirements for detecting change and design programs that take advantage of the strengths of VLSA-based monitoring and account for potential challenges.

Technologies and techniques for acquiring, processing, and extracting information from VLSA images are rapidly evolving. For this reason, it is of paramount importance to understand the statistical implications of designing and implementing a rangeland monitoring and assessment program based on VLSA imagery and the aspects of designing VLSA image studies that are different than traditional field-based studies, such as image scale, sample design limitations, and the need for validation of estimates. The extent to which VLSA imagery will be useful as a tool for understanding the status and trend of rangelands depends as much on the ability to build the imagery into robust programs as it does on the ability to quickly and relatively easily collect VLSA images over large landscapes.

LITERATURE CITED

Addicott, J. F., J. M. Aho, M. F. Antolin, D. K. Padilla, J. S. Richardson, and D. A. Soluk. 1987. Ecological neighborhoods: scaling environmental patterns. *Oikos* 49:340–346.

Bethlehem, J. G. 2009. Applied survey methods: a statistical perspective. Hoboken, NJ, USA: John Wiley & Sons.

Booth, T. D., and S. E. Cox. 2008. Image-based monitoring to measure ecological change in rangeland. *Frontiers of Ecology and the Environment* 6:185–190.

Booth, T. D., S. E. Cox, and R. D. Berryman. 2006. Point sampling digital imagery with "Samplepoint." *Environmental Monitoring and Assessment* 123:97–108.

Booth, T. D., S. E. Cox, C. Fifield, M. Phillips, and N. Williamson. 2005a. Image analysis compared with other methods for measuring ground cover. *Arid Land Research and Management* 19:91–100.

Booth, T. D., S. E. Cox, and D. E. Johnson. 2005b. Detection-threshold calibration and other factors influencing digital measurements of ground cover. *Rangeland Ecology & Management* 58:598–604.

Booth, T. D., S. E. Cox, M. Louhaichi, and D. E. Johnson. 2004. Technical note: lightweight camera stand for close-to-earth remote sensing. *Journal of Range Management* 57:675–678.

Booth, T. D., D. Glenn, B. Keating, J. Nance, S. E. Cox, and J. P. Barriere. 2003. Monitoring rangeland watersheds with very-large scale aerial imagery. First Interagency Conference on Research in the Watersheds; 27–30 October 2003; Benson, AZ, USA. Washington, DC, USA: USDA-ARS. 753 p.

Booth, T. D., and P. T. Tueller. 2003. Rangeland monitoring using remote sensing. *Arid Land Research and Management* 17:455–467.

Breidt, F. J., and W. A. Fuller. 1999. Design of supplemented panel surveys with application to the National Resources Inventory. *Journal of Agricultural, Biological, and Environmental Statistics* 4:391–403.

Cochran, W. G. 1977. Sampling techniques. 3rd ed. New York, NY, USA: John Wiley & Sons.

Duniway, M. C., J. W. Karl, S. Schrader, N. Baquera, and J. E. Herrick. 2012. Rangeland and pasture monitoring: an approach to interpretation of high-resolution imagery focused on observer calibration for repeatability. *Environmental Monitoring and Assessment* 184:3789–3804. doi:10.1007/s10661-011-2224-2.

Elzinga, C. L., D. W. Salzer, and J. W. Willoughby. 1998. Measuring and monitoring plant populations. Denver, CO, USA: US Department of the Interior–Bureau of Land Management National Applied Resource Sciences Center. 477.

FARINA, A. 1998. Principles and methods in landscape ecology. London, UK: Chapman & Hall. 412 p.

FULLER, W. A. 1999. Estimation procedures for the United States National Resource Inventory. *In:* Proceedings of the Survey Methods Section, Statistical Society of Canada Annual Meeting; 6–9 June 1999; Regina, SK, Canada. Ottawa, ON, Canada: Statistical Society of Canada. p. 39–44.

GODINEZ-ALVAREZ, H., J. E. HERRICK, M. MATTOCKS, D. TOLEDO, AND J. VAN ZEE. 2009. Comparison of three vegetation monitoring methods: their relative utility for ecological assessment and monitoring. *Ecological Indicators* 9:1001–1008.

GREGOIRE, T. G., AND H. T. VALENTINE. 2008. Sampling techniques for natural and environmental resources. Boca Raton, FL, USA: Taylor & Francis. 496 p.

HERRICK, J. E., J. W. VAN ZEE, K. M. HAVSTAD, L. M. BURKETT, AND W. G. WHITFORD. 2009. Monitoring manual for grassland, shrubland, and savanna ecosystems. Las Cruces, NM, USA: USDA-ARS Jornada Experimental Range. 36 p.

HUNT, E. R., JR., J. H. EVERITT, J. C. RITCHIE, M. S. MORAN, T. D. BOOTH, G. L. ANDERSON, P. E. CLARK, AND M. S. SEYFRIED. 2003. Applications and research using remote sensing for rangeland management. *Photogrammetric Engineering and Remote Sensing* 69:675–693.

KARL, J. W., M. C. DUNIWAY, AND T. S. SCHRADER. 2011 A technique for estimating rangeland canopy-gap size distributions from high-resolution digital imagery. *Rangeland Ecology & Management* 65:196–207.

KOTLAIR, N. B., AND J. A. WIENS. 1990. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. *Oikos* 59:253–260.

LALIBERTE, A. S., J. E. HERRICK, A. RANGO, AND C. WINTERS. 2010. Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. *Photogrammetric Engineering and Remote Sensing* 76:661–672.

LALIBERTE, A. S., A. RANGO, AND E. L. FREDRICKSON. 2006. Separating green and senescent vegetation in very high resolution photography using an intensity-hue-saturation transformation and object based classification. *In*: Proceedings of the American Society for Photogrammetry and Remote Sensing Annual Conference; 1–5 May 2006; Reno, NV, USA. CD-ROM.

LALIBERTE, A. S., A. RANGO, K. M. HAVSTAD, J. F. PARIS, R. F. BECK, R. MCNEELY, AND A. L. GONZALEZ. 2004. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of the Environment* 93:198–210.

LILLESAND, T. M., AND R. W. KIEFER. 1994. Remote sensing and image interpretation. New York, NY, USA: John Wiley & Sons. 750 p.

LOHR, S. L. 2009. Sampling: design and analysis. 2nd ed. Boston, MA, USA: Duxbury Press. 608 p.

LUSCIER, J. D., W. L. THOMPSON, J. M. WILSON, B. E. GORHAM, AND L. D. DRAGUT. 2006. Using digital photographs and object-based image analysis to estimate percent ground cover in vegetation plots. *Frontiers of Ecology and the Environment* 4:408–413.

MACKINNON, W. C., J. W. KARL, G. R. TOEVS, J. J. TAYLOR, M. S. KARL, C. S. SPURRIER, AND J. E. HERRICK. 2011. BLM core terrestrial indicators and methods. Tech Note 440. Denver, CO, USA: US Department of the Interior–Bureau of Land Management National Operations Center. 13 p.

MARSETT, R. C., J. QI, P. HEILMAN, S. H. BEIDENBENDER, M. C. WATSON, S. AMER, M. WELTZ, D. GOODRICH, AND R. MARSETT. 2006. Remote sensing for grassland management in the arid southwest. *Rangeland Ecology & Management* 59:530–540.

NATIONAL RESEARCH COUNCIL. 1994. Rangeland health: new methods to classify, inventory, and monitor rangelands. Washington, DC, USA: National Academy Press. 182 p.

NUSSER, S. M., F. J. BREIDT, AND W. A. FULLER. 1998. Design and estimation for investigating the dynamics of natural resources. *Ecological Applications* 8:234–245.

NUSSER, S. M., AND J. J. GOEBEL. 1997. The National Resources Inventory: a long-term multi-resource monitoring programme. *Environmental and Ecological Statistics* 4:181–204.

RANGO, A., A. S. LALIBERTE, J. E. HERRICK, C. WINTERS, K. M. HAVSTAD, C. STEELE, AND D. BROWNING. 2009. Unmanned aerial vehicle based remote sensing for rangeland assessment, monitoring and management. *Journal of Applied Remote Sensing* 2:033542.

SARNDAL, E., B. SWENSSON, AND J. WRETMAN. 2003. Model assisted survey sampling. New York, NY, USA: Springer-Verlag. 709 p.

STRAHLER, A. H., C. WOODCOCK, AND J. A. SMITH. 1986. On the nature of models in remote sensing. *Remote Sensing of the Environment* 20:121–139.

STRAND, E., A. M. S. SMITH, S. C. BUNTING, L. VIERLING, D. B. HANN, AND P. E. GESSLER. 2006. Eavelet estimation of plant spatial patterns in multi-temporal aerial photography. *International Journal of Remote Sensing* 27:2049–2054.

THOMPSON, S. K. 2012. Sampling. 3rd ed. New York, NY, USA: John Wiley & Sons. 472 p.

TURNER, M. G., R. V. O'NEILL, R. H. GARDNER, AND B. T. MILNE. 1989. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology* 3:153–162.

USDA-NRCS. 2009. National resources inventory grazing land on-site study: handbook of instructions. Washington, DC, USA: USDA-NRCS. Available at: http://www.nrisurvey.org/nrcs/Grazingland/2010/instructions/instruction.htm. Accessed 4 June 2012.

WEBER, K. T., J. THEAU, AND K. SERR. 2008. Effect of co-registration error on patchy target detection using high-resolution imagery. *Remote Sensing of Environment* 112:845–850.

WOODCOCK, C., AND A. H. STRAHLER. 1987. The factor of scale in remote sensing. *Remote Sensing of the Environment* 21:311–332.