Design and Analysis of Rangeland Experiments Along Continuous Gradients

Nicola Koper,[1] *Darcy C. Henderson,*[2] *John F. Wilmshurst,*[3] *Patrick J. Fargey,*[4] *and Robert A. Sissons*[5]

Authors are [1]*Assistant Professor, Natural Resources Institute, University of Manitoba, Winnipeg, MB R3T 2N2, Canada;* [2]*Visiting Fellow and* [3]*Grassland Ecologist, Parks Canada, 145 McDermot Ave, Winnipeg, MB R3B 0R9, Canada; and* [4]*Species at Risk Biologist and* [5]*Conservation Biologist, Grasslands National Park of Canada, Box 150, Val Marie, SK S0N 2T0, Canada.*

Abstract

Measuring the effects of grazing intensity on the structure, composition, and function of grassland ecosystems has been a perennial challenge. Space limits replication, few replicates limit statistical power, and categorical treatments limit interpretation of effects among treatment levels. Treating grazing as a continuous rather than categorical variable can permit large-scale experiments to be conducted with fewer constraints on treatment replication to maintain statistical power. Using power analysis on a grazing experiment recently initiated in Grasslands National Park of Canada, we demonstrate that the continuous approach permits the use of fewer pastures, while maintaining the large pasture size required to allow realistic grazing behavior by cattle and improving our ability to answer biologically relevant questions regarding grazing effects on grassland ecosystems. We contend that this approach, when applied to grazing experiments, will help test hypotheses related to how grassland ecosystems respond to a gradient of disturbance regimes.

Resumen

El medir el efecto de la intensidad de pastoreo en la estructura, composición, y funcionamiento del ecosistema del pastizal ha sido un reto permanente. Las limitantes de espacio para las repeticiones, así como pocas repeticiones limitan el poder estadístico, y los tratamientos categóricos limitan la interpretación de los efectos entre los niveles de los tratamientos. El utilizar el tratamiento del pastoreo como una variable contínua en lugar de variable categórica puede hacer posible experimentos a larga escala con pocas restricciones en la repetición de los tratamientos para poder mantener el poder estadístico. Utilizando análisis de poder en un experimento recién iniciado en los pastizales de un parque nacional de Canadá, demostramos que el utilizar el pastoreo como una variable continúa permite el uso de menos potreros, al mismo tiempo que se mantuvo un potrero de tamaño grande que permitiera un comportamiento normal por el ganado y a la vez mejorar nuestra habilidad para contestar las preguntas biológicas relevantes a los efectos del pastoreo en los ecosistemas de pastizales. Sugerimos que este método, al aplicarse a los experimentos de pastoreo, ayudará a probar la hipótesis a como los ecosistemas de pastizales pueden responder al grado de disturbio.

INTRODUCTION

One of the principal objectives of research in rangelands has been to study the effects of grazing on the structure and function of ecological communities (e.g., Biondini et al. 1998; Fuhlendorf and Engle 2001). An important goal has been to develop an understanding of how equilibrium concepts of rangeland succession relate to nonequilibrium models across temporal and spatial scales (Briske et al. 2003). However, experimental approaches to test these hypotheses have not kept pace with the changes in the theories. Our objective here is to compare two sampling designs that could be used to examine a continuum of responses of grassland floral and faunal communities to a gradient of grazing intensity in a controlled field experiment.

Two hypotheses dominate thinking on grazing effects on grassland community structure and function (Briske et al. 2003). The range model predicts that grazing effects are continuous and reversible (Milchunas et al. 1988). Alternatively, the states and transition model predicts that disturbance can result in relatively sudden changes, and that these changes might not be reversible (Laycock 1991), in effect, creating an equilibrium state, at least temporarily. However, within the domain of attraction of an equilibrium state, disturbance effects can be reversible, suggesting that the range model is a special case of the states and transitions model (Briske et al. 2003). Hence, we can expect that, within a community, a gradient of grazing intensities can elicit a gradient of ecosystem responses (e.g., HilleRisLambers et al. 2001; Jonas and Joern 2007).

Ecological responses to grazing are not only sensitive to the grassland's history of grazing or to its climatic regime, but also to the spatial and temporal scale at which the disturbance has occurred (Briske et al. 2005). Cattle behavior (Howery et al. 1998) and ecosystem responses to grazing pressure (Rietkerk et al. 2000) can vary considerably at large compared to small scales, even with similar disturbance intensities (Fuhlendorf and Smeins 1999). In larger pastures, habitat and vegetation

Table 1. Characteristics of manipulated grazing experiments in North American mixed-grass and short-grass prairie, 1916 to present.

Location	Initiated (Ceased)	Treatments	n[1]	Area (ha)	References
Wyoming "High Plains Experiment Station"	1982	Exclusion	2	~2.5	Hart et al. 1988
		17% utilization	2	~64.0	Schuman et al. 1999
		33% utilization	2	~12.0	
		50% utilization	2	~9.0	
North Dakota "Mandan"	1916	Exclusion	1	< 0.1	Frank et al. 1995
		Moderate utilization	1	46.0	
		High utilization	1	16.0	
North Dakota "Streeter"	1979	Exclusion	3	13.2	Biondini et al. 1998
	1988	45% utilization	3	13.2	
	1988	77% utilization	3	13.2	
Colorado "Central Plains Experiment Station"	1939	Exclusion	3	2.0	Milchunas et al. 1989
		Light utilization	1	138.0	Hart 2001
		Moderate utilization	1	138.0	
		High utilization	1	138.0	
Oklahoma "Clinton"	1990 (1996?)	Exclusion	0	N/A[2]	Gillen et al. 2000
		Moderate to high utilization range	6	65.0	
Texas "Sonora"	1948	Exclusion	2	12.0	Fuhlendorf and Smeins 1997
		Moderate utilization	4	24.0	
		High utilization	2	32.0	
Saskatchewan "Swift Current"	1989	May–June	2	1.6	Schellenberg et al. 1999
		July–August	2	1.6	
		August–September	2	1.6	
		September–October	2	1.6	
Alberta "Onefour Blacktail"	2001	Exclusion	6	0.3	W. Willms, personal communication, February 2006
		June–August	3	265.0	
		September–November	3	265.0	
Alberta "Onefour Sheep"	1950 (1970)	Exclusion	1	< 0.1	Smoliak et al. 1972
		Light utilization	1	~16.0	
		Moderate utilization	1	~16.0	
		High utilization	1	~16.0	

[1]Number of experimental units.
[2]N/A indicates that no pastures excluded cattle grazing.

structure are typically more heterogeneous than in smaller pastures. Further, cattle graze more selectively in larger pastures, further increasing heterogeneity (Hart et al. 1993). An understanding of the effects of commercial grazing systems on prairie ecosystems therefore requires studies within appropriately sized, relatively large pastures. Using a hierarchical and nested sampling approach enables a comparison of the effects of grazing at multiple spatial scales, a critical objective for modern grazing studies (Briske et al. 2005).

We argue that evaluating ecosystem responses to a grazing intensity gradient at spatial scales relevant for range managers requires a design that incorporates both large experimental units (pastures) and a gradient of grazing intensity treatments. This is clearly challenging, and thus it is not surprising how few grazing-intensity experiments have been conducted at scales larger than a single range site, using actual livestock (as opposed to mechanical defoliation), or with more than three grazing intensities (Table 1).

Our literature search suggests that in North America, no long-term manipulative grazing experiments in low productivity grasslands have incorporated a sufficient range of grazing intensities to describe the linear or nonlinear effect of grazing intensity, which is a key characteristic of ecosystem responses to grazing (Milchunas et al. 1988). The closest example is Gillen et al. (2000), who imposed a range of grazing intensities on six 65-ha pastures, but included no ungrazed pastures in their experimental design. Previous experiments have not incorporated more than 30 ha (Biondini et al. 1998) of ungrazed prairie, which is below the minimum patch size for many grassland birds (Davis 2004). Only two previous studies have used pastures of more than 100 ha in size (Hart 2001), and all are smaller than commercial pastures (e.g., in southern Alberta, mean = 456 ha, SD = 632 ha; Henderson 2005). Further, none of these studies have incorporated a before-treatment sampling period, which often is critical for linking community responses to grazing disturbance (Underwood 1994). In general, grazing experiments have been dominated by discontinuous treatments with few treatment levels, little spatial replication, and small pastures.

Here we present a grazing intensity experiment in northern mixed-grass prairie that we argue addresses these problems. The objective of the experiment is to understand effects of
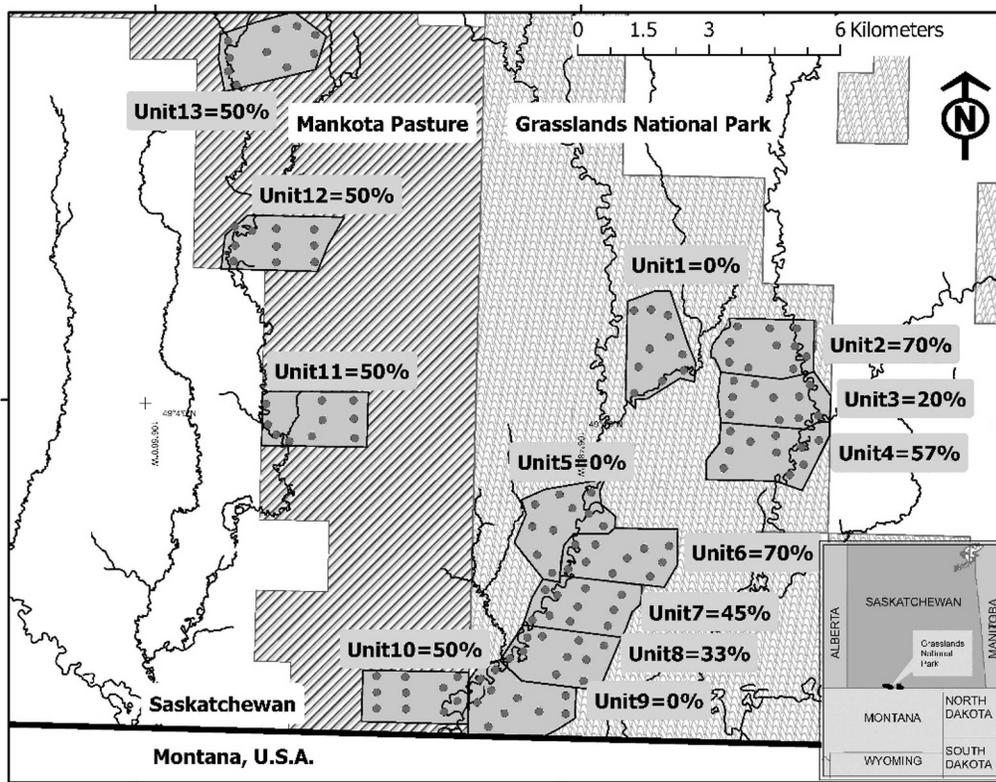
Figure 1. Schematic of experimental layout for the grazing experiment in and around Grasslands National Park of Canada, Saskatchewan. Experimental units (pastures) are polygons shaded dark gray with sampling plots indicated by dark gray dots. By each experimental unit is a label indicating its treatment under the continuous experimental design (target percent offtake). The area within the park boundaries is shaded light gray, and Mankota community pastures are shaded in diagonal stripes.

grazing intensity on multiscale heterogeneity and community structure in mixed-grass prairie communities. We designed the experiment to be conducted at large spatial scales, incorporate pretreatment sampling, and impose grazing treatments along a gradient of intensities. The resulting experimental design provides a case study for discussing the problems of some past experiments and introducing an analytical approach not typically applied in rangeland research experiments. Our objective in the current paper was to evaluate the relative power of a traditional, categorical design, with a design in which grazing intensity is treated as a continuous variable. To accomplish the required comparison of power in continuous relative to categorical designs, we developed a novel approach for conducting a priori power analyses for both regression and analysis of variance (ANOVA) analyses, when preliminary data on the gradient are absent.

METHODS

Study Area

The experiment ("Biodiversity and Grazing Experiment") was designed for the East Block of Grasslands National Park of Canada (GNPC), located in southern Saskatchewan (located at approximately lat 49°01′00″N, long 106°49′00″W; Fig. 1). This area is characteristic of northern mixed-grass prairie, and supports a wide range of landscapes varying in hydrologic, soil, and vegetation characteristics. The park is in a subhumid climate zone, with mean annual precipitation of approximately 350 mm and potential annual evapotranspiration of approximately 347 mm (Kottek et al. 2006).

The experimental area occupies 26.5 km^2 in the East Block, and four additional pastures in the adjacent Mankota Community Pastures will function as long-term grazed control pastures for the study (Fig. 1). All pastures are characterized by a relatively low relief and elevation (750 to 850 m above sea level) landscape of glacial till and alluvial deposits. Three broad vegetation–landscape units occur in the experimental area: riparian shrublands, upland grasslands, and valley grasslands (Michalsky and Ellis 1994). This area was never cultivated, nor heavily grazed by livestock in the time between homesteading in the early 1900s and purchase by Parks Canada, and there are few cross-fences. The experimental area has been ungrazed since approximately 1992, whereas the Mankota Community Pastures continued to be grazed annually throughout that period.

Within the experimental area, nine pastures have been constructed (Fig. 1). They average 296 ha, and vary from this by less than 8% (SD = 14, range = 280–321 ha). Pretreatment baseline sampling was conducted May–August 2006 and 2007. In June 2008, yearling steers were introduced to six of the GNPC pastures at stocking rates predicted to result in average annual forage utilization rates ranging from 20% to 70% (very low to high utilization). The remaining three GNPC pastures will be ungrazed controls. All pastures were configured to ensure that they were similar in 1) shape and size; 2) proportion

of lowland, slope, and upland habitat; 3) location of a natural water source; and 4) plant communities. All pastures will also include an anthropogenic upland water source, to be consistent with the regional pasture management. Within each pasture, 10 plots will be surveyed: four in the valley grasslands, and six in the upland grasslands. Here, we treat the pastures as the units of replication ($n = 9$ experimental pastures).

Study Design

Our experimental objective lent itself to a before–after, control–impact (BACI) design, which measures variation among the pastures prior to manipulations, and both spatial and temporal variation during the experimental manipulation (Osenberg et al. 1994). This design controls for both spatial and temporal variability among replicates. Our preliminary analysis evaluated the degree to which pre-existing environmental patterns in our response variables occur in our study area. Our study went a step further to incorporating a "beyond BACI" design (Underwood 1994), which, in addition to the BACI component, includes multiple treatment levels in the impact component.

When space is limiting, there is a significant tradeoff between establishing numerous treatments that are simultaneously replicated in space, and establishing treatments of appropriately large spatial scale relative to the phenomena under question (Fisher 2000). Our interest in maintaining relatively large experimental units meant that opportunities for replication of all treatments were limited, because only nine pastures of approximately 300 ha could be accommodated in our study area. An experimental design that treated grazing intensity as a categorical variable and required replicates within each level of grazing intensity would restrict our grazing intensities to ungrazed, light (such as 35% utilization), and high (such as 70% utilization), each with only three replicate pastures. For simplicity we refer to this as a categorical design, and would anticipate that it would be analyzed using a statistical approach such as ANOVA.

Alternatively, however, grazing can be treated as a continuous variable (the "continuous design"), which would typically be analyzed using a statistical approach such as regression. In this case, we would replicate grazing intensities only at the ungrazed and maximally grazed intensities. Three ungrazed pastures would be included in the design, to provide high confidence in our estimate of environmental parameters in the absence of grazing. Two pastures would be grazed at the maximum grazing intensity of 70% utilization, to minimize the likelihood of influential outliers having a strong effect on the regression slope (e.g., Quinn and Keough 2002). Intermediate grazing intensities (20, 33, 45, 57% removal) would each be applied to a single pasture. This design would enable us to determine whether relationships between grazing intensity and various dependent variables are linear or curvilinear. Pastures were assigned randomly to each grazing intensity a priori.[1]

---

[1] For the implementation of the study in 2008 and after the acceptance of this manuscript, stocking rates had to be changed to the following to accommodate stream and riparian research added to the protocol: Unit 2 = 20%, Unit 3 = 57%, Unit 4 = 70%, Unit 6 = 33%, Unit 8 = 70%. All simulations for the power analyses in this manuscript were conducted assuming stocking rates as in Figure 1.

Vegetation and Ground Squirrel Burrow Surveys

At each of the 10 survey plots per pasture, a 1000-m^2 (50 m × 20 m) Modified Whittaker sampling plot was established to measure plant community richness (Stohlgren et al. 1995) between June and August. A walk-through survey identified vegetation species presence, as well as the presence of microsites such as burrows (ground squirrel hole density). A foliar coverage class method was used in 10 1 m × 0.5 m frames to visually estimate vascular plant species' relative abundance. In this study, these were summed across frames to determine coverage per Whittaker plot. Visual estimates of bare ground, lichen, black algae, and litter were also taken within the 10 frames. Vegetation heights and litter depths were measured using a meter stick, and averaged among frames. Vegetation was clipped from within four, 1 m × 1 m frames to determine biomass.

Avian Surveys

We estimated avian densities using 10 5-min, 100-m radius point counts per pasture, in which all songbirds seen or heard within plots were identified and recorded. Vegetation plots were nested within point-count plots. Surveys were conducted between dawn and 0950 hours, in the absence of precipitation and winds exceeding 20 km · hr^{-1}. Each point-count plot was surveyed three times between 25 May and 20 June. Before analysis, results per species were averaged across rounds to estimate species densities within plots.

Power Analyses

We evaluated power at the scale of the pasture, because this is the spatial scale at which we have the fewest replicates and therefore the least power. Pasture-level effects for both avian and vegetation surveys were estimated by averaging results across all plots within each pasture, from 2006. We calculated effect sizes from this study or from the literature (see below) and we set the significance threshold (α) to 0.1, and power (1-β) of at least 0.8. We conducted a priori power analyses using a number of dependent variables that we expected to be sensitive to grazing. For vegetation indicators, we evaluated power using one increaser (blue grama [*Bouteloua gracilis* (H.B.K.) Lag. ex Seud.]), one decreaser (northern wheatgrass [*Elymus lanceolatus* (Scribn. and Smith) Gould]), and one measure of vegetation structure (vegetation height). For vertebrate indicators, we evaluated power using one species that prefers shorter vegetation (chestnut-collared longspur [*Calcarius ornatus*]), one species that prefers taller vegetation (Savannah sparrow [*Passerculus sandwichensis*]), and an index of ground squirrel abundance (sum of number of Richardson's [*Spermophilus richardsonii*] and thirteen-lined ground squirrel [*Spermophilus tridecemlineatus*] burrows).

Chestnut-collared longspurs can respond positively to grazing pressure, or can select moderately grazed fields but avoid heavily grazed fields (Milchunas et al. 1998). We therefore evaluated our power to detect both linear and quadratic (curvilinear) effects of grazing on chestnut-collared longspur densities.

Effect Size Estimation. One of the challenges of a priori power analysis is the need to forecast the size of treatment effects before any measurements have been made. It is clear from the

literature that intense grazing can result in a range of possible effect sizes among different studies, even when direction of the effect can be confidently predicted. Because we wanted to predict which design would be more powerful under a range of possible responses, we tested power under conditions of both relatively large and small effect sizes.

To predict the effect size of grazing on response variables of interest, we used 1) effect sizes recorded in the literature from other grazing studies conducted in mixed-grass prairies, 2) effect sizes calculated from a study conducted in a mixed-grass prairie in Alberta between 2000 and 2002 (Koper and Schmiegelow 2006), or 3) compared the response in the ungrazed sites and grazed sites, surveyed in our study area in 2006 as part of the pretreatment sampling for our grazing experiment. For example, if we found that chestnut-collared longspur density in grazed fields was 270% of the density in ungrazed fields, the effect size was 2.7 times the density in ungrazed fields (the proportional effect, or PE). Where data were available, we preferentially used methods 1 or 2 to calculate effect sizes, because the light grazing intensity in the grazed pastures (24% removal, T. Teetaert, unpublished data, 2006) was likely to lead to an underestimate of the predicted effects of grazing.

We used the smallest and largest effect size found in our analyses or literature search to represent the minimum and maximum effect size we would be likely to detect, respectively. For our variable measuring ground squirrel hole density, the only available data were from the present study. In this case, we considered the observed difference between grazed and ungrazed sites to represent a biologically small difference. Grazed pastures are managed with the intention of allowing no more than 50% of the vegetation to be removed annually, although environmental variability can make it difficult to achieve this goal exactly. Our highest grazing intensity is intended to result in the removal of 70% of the vegetation produced annually. We therefore calculated the biologically large effect by multiplying the observed difference between grazed and ungrazed sites by 70/50.

Data Simulation Using Calculated Effect Sizes. Where linear effects of grazing intensity were predicted, we simulated the effect of grazing in the pastures that will have cattle introduced at the maximum stocking rate of 70% utilization by adding the predicted effects of grazing to each of the observed pretreatment field values of the response variable. This introduced the predicted effect of grazing while maintaining the underlying variability in our data. This was repeated for both maximum and minimum predicted effect sizes.

To calculate the effect of intermediate grazing intensities for the continuous design, we multiplied the maximum predicted effect of grazing by 20/70, 33/70, 45/70, and 57/70, relative to the grazing intensity to be imposed on each pasture, and added this value to the response value for each respective pasture. To calculate the effect of the grazing intensity for the categorical design, we first assigned the pasture intended to get the 57% utilization treatment to the 70% utilization treatment, to bring the number of replicates within each of the control and high-intensity grazing treatments to three. We then multiplied the maximum predicted effect of grazing by 35/70 and added this to the response values of the three pastures assigned to the intermediate grazing intensity. A similar approach was used to estimate quadratic effects for both the categorical and continuous designs, but in this case, the maximum effect was assigned to an intermediate grazing intensity (57%, Milchunas et al. 1998).

For example, to calculate the predicted effect of grazing within the 57% treatment, which we plan to assign to pasture 4, on chestnut-collared longspurs, assuming a linear effect of grazing, we used the following equation:

$$OD_4 + OD_4 \times 57/70 \times PE \qquad [1]$$

where OD_4 is the average of the observed density of chestnut-collared longspurs in point-count plots within pasture 4 in 2006, and PE is the proportional effect of grazing as calculated from the literature (Biondini et al. 1998).

Power Analysis Calculations. Effect sizes were first calculated from the predicted effects of grazing relative to the underlying variance in the empirical data, using means and SD for the categorical analyses, and R^2 values for the continuous analyses. We calculated means and SD for the categorical data sets in Microsoft Excel 2003. We calculated R^2 values for the continuous data sets in S-plus 6.2. We then used G*power 3.0.8 (Faul et al. 2007) to calculate the sample sizes required to achieve at least 0.8 power with $\alpha = 0.1$ and the appropriate effect size, specific to each response variable.

RESULTS

The continuous design was more powerful than the categorical design for all of the large effect sizes (Table 2), and for four out of seven small effect sizes. When the continuous design was more powerful than the categorical design, the sample size required for the categorical design ranged from 20% to 247% larger than the sample size required for the continuous design (Table 2). For northern wheatgrass, the predicted required sample size was marginally higher for the continuous design for detecting a small effect (13 vs. 12), whereas the predicted required sample size for detecting a large effect was substantially higher for the categorical design (6 vs. 3), again suggesting that in general, the continuous design was more powerful than the categorical design for detecting effects of grazing on northern wheatgrass.

Our sample size would clearly be inadequate for detecting small effects of grazing on Savannah sparrows (Table 2) using either the continuous or categorical design. It is also unlikely that we would detect large effects of grazing on Savannah sparrows, although our chances of doing so would be substantially better with the continuous design as opposed to the categorical design. This result arises because of the high variability in Savannah sparrow densities among pastures, and because their densities appear to be relatively insensitive to grazing (Fig. 2; Table 2).

Initially, the categorical design appeared to be more powerful in detecting grazing effects on densities of ground squirrel burrows (Table 2). However, this occurred because pretreatment densities of ground squirrel burrows follow a spurious nonlinear pattern (Fig. 3a). When a small linear effect of grazing was imposed on these data, the original spurious trend resulted in an apparently nonlinear increasing trend for the

Table 2. Predicted sample sizes to test for change in grassland community variables relative to predicted small and large grazing effects in southwestern Saskatchewan. All literature estimates of effect of intense grazing on vegetation height were 44%, so we conducted our power analysis only at this effect size.

Response Variable	Function	Effect Size					
		Small			Large		
		% change	Sample size continuous	Sample size categorical	% change	Sample size continuous	Sample size categorical
Blue grama cover[1]	linear	+22[2]	30	36	+60[2,3]	4	6
Vegetation height	linear	−44[4]	4	6	N/A[5]	—	—
Northern wheatgrass cover	linear	−43[6]	13	12	−94[2]	3	6
Ground squirrel hole density	linear	+24[7]	12	9	+34[7]	8	9
Chestnut-collared longspur density	linear	+69[6]	15	24	+270[8]	5	6
Chestnut-collared longspur density	quadratic[8]	+69[6]	9	21	+270[8]	6	9
Savannah sparrow density	linear	−17.6[9]	2 056	144	−100[10]	17	42

[1]Valley grassland community type. All other response variables in upland grassland community type.
[2]Milchunas et al. 1998.
[3]Frank et al. 1995.
[4]Bai et al. 2001.
[5]N/A indicates that only one analysis was done using vegetation height, because all literature estimates of effect of intense grazing on vegetation height were 44%.
[6]Biondini et al.1998.
[7]Mankota Community Pasture, this study.
[8]Model includes an additional parameter.
[9]Koper and Schmiegelow 2006.
[10]Dale 1983.

continuous and categorical designs (Figs. 3b and 3c). However, both patterns are misleading. To demonstrate this, we conducted the power analysis for the categorical design using the unaltered ground squirrel data, with no effect of grazing added. We again calculated that we would require nine pastures to achieve sufficient power to detect differences among treatments, despite the fact that there should theoretically be no intrinsic differences among the groups before grazing is imposed. The low sample size apparently required for the categorical design therefore reflects spurious differences among the treatment groups.

The sample size in our study is sufficiently large to detect effects of biologically small effects for two of the six response variables that we examined (assuming a quadratic response by chestnut-collared longspurs), and can detect a biologically large effect for five of the six response variables (Table 2).

DISCUSSION

The continuous design, analyzed using linear regression, was consistently more powerful than the categorical design, analyzed using ANOVA, for what we contend are three key reasons. First, the continuous design makes use of the pattern in the order of responses to determine significance. The categorical design compares the size of the differences among groups relative to the variance within groups, without regard to the order among those groups. Ignoring the order of the groups results in the loss of potentially useful information (e.g., Cingolani et al. 2005; Sasaki et al. 2008). Second, the continuous design uses all of the data to estimate the variance or error around the regression line, resulting in a more accurate and usually smaller estimate of the variance. In contrast, in the categorical design, the within-treatment variance is estimated from only those points within each treatment. This typically

results in low power to detect differences among treatments. Third, the categorical design requires equal sample sizes among groups to maximize power, whereas the regression design does not have this restriction (Quinn and Keough 2002).

Perhaps of greater interest to rangeland managers is that the continuous design lends itself to interpolation. This means that the effects of the independent variable at points on the curve that are not explicitly sampled can be interpreted from the other points by using the information from the shape of the response curve. This can be a significant benefit over a categorical design, in which all inference is restricted to the levels of the treatments imposed (Quinn and Keough 2002). Interpolation can be critical for managers who need to predict the grazing intensity that would achieve a target density for a focal species. Although a replicated categorical design of ungrazed, lightly grazed, and heavily grazed pastures could be analyzed using linear regression, interpolation cannot be trusted with that design, because the shape of the curve is defined by only three points.
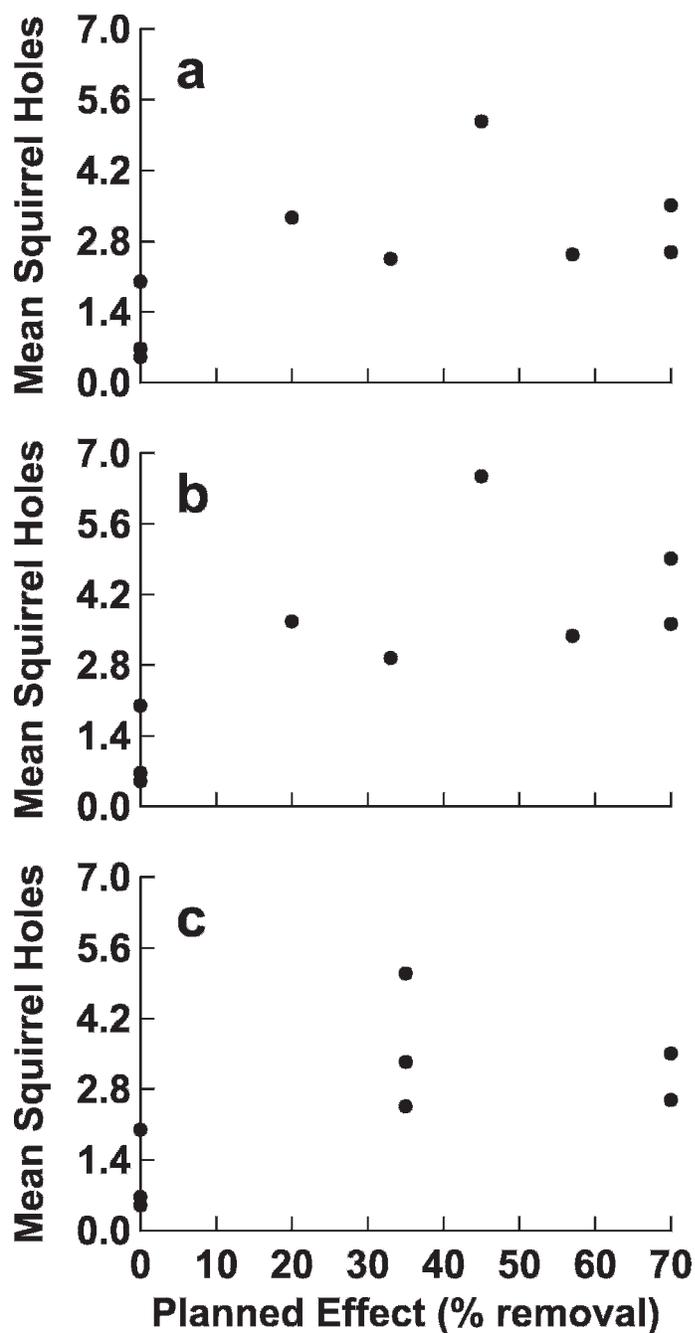
Although the continuous design was generally more efficient and powerful, the exceptions require a more detailed review. Detecting significant change in ground squirrel burrow density with small grazing effects required only nine samples in a categorical design, but 12 in the continuous design. However, the sample size for the categorical design is influenced by a spurious trend in the data. The weak, nonlinear pattern that existed in the experimental units prior to any grazing treatment being imposed is an artifact of the spatial ordering of our experimental units unrelated to grazing, which can be a hazard of random assignment of treatments (Krebs 1989). This strongly emphasizes the need for sampling pastures before grazing regimes are introduced (Underwood 1994). Such spurious effects can explain inconsistent results among previous studies (e.g., cf. Schuman et al. 1999; Romo and Bai 2004). Conducting pretreatment sampling is critical, because no statistical analysis
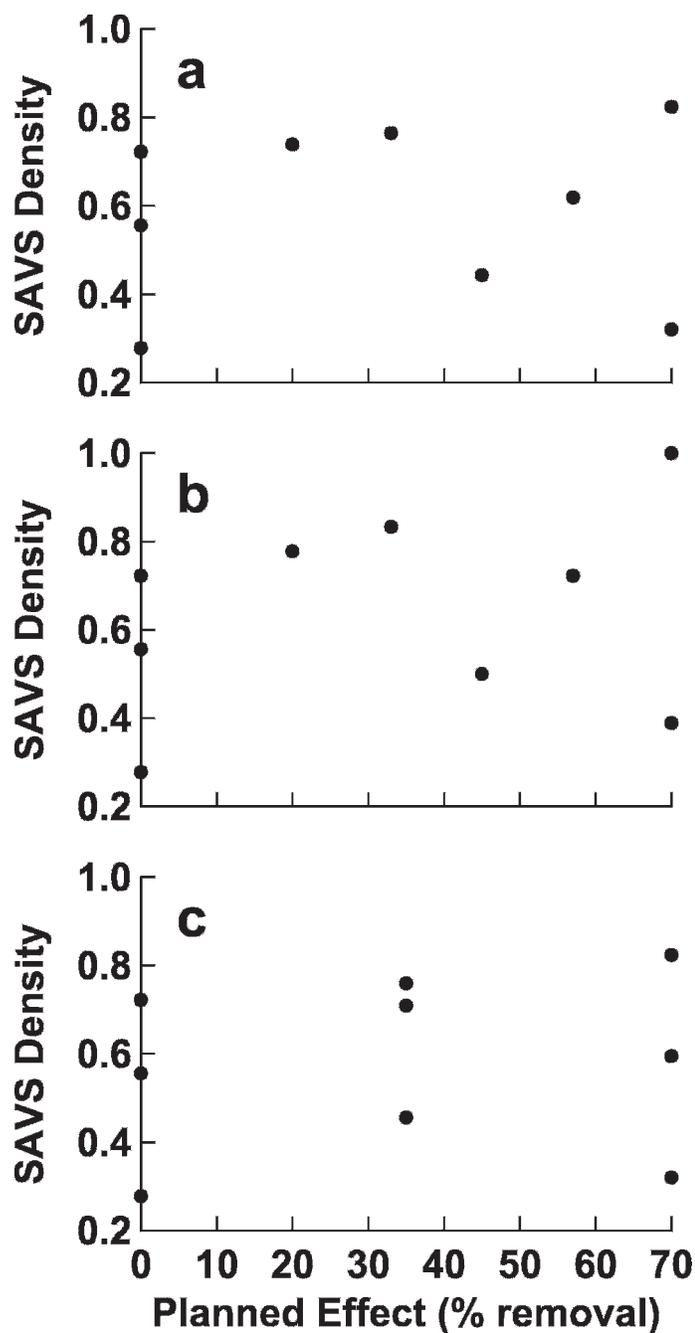
Figure 2. Average density of Savannah sparrows (SAVS) per 3.2-ha point-count plot per pasture in Grasslands National Park of Canada in southern Saskatchewan in 2006, relative to the target percent offtake to be experimentally removed by cattle per pasture 2008–2017. **a**, 2 yr before introduction of cattle to the grazing experiment. **b**, Predicted densities following continuous design with simulated linear response. **c**, Predicted densities following categorical design with simulated linear response.

Figure 3. Average density of ground squirrel burrows per 1 000-m^2 Whittaker plot per pasture in Grasslands National Park of Canada in southern Saskatchewan in 2006, relative to the target percent offtake to be experimentally removed by cattle per pasture, 2008–2017. **a**, 2 yr before introduction of cattle to the grazing experiment. **b**, Predicted densities following continuous design with simulated linear response. **c**, Predicted densities following categorical design with simulated linear response.

can remove the effect of spurious patterns intrinsic to a data set, if the presence of those patterns is undetected.

If burrow density was the only response variable that we were interested in, then the best approach to dealing with this bias would be to reassign the grazing treatments to pastures to remove the trend. However, we are interested in several variables, and initial attempts to remove the pattern by reassigning pastures to grazing treatments introduced spurious patterns to other variables (N. Koper, unpublished data, 2007). We concluded that burrow density is not of sufficiently high priority for us to reassign the grazing treatments, and we will instead depend on measures indicating change over time to determine whether there are effects of grazing intensity on burrow densities.

A second message that emerges from the power analyses is that we might have insufficient power to detect some of the expected grazing effects with either the continuous or categorical design. This is highlighted by Savannah sparrow densities, which we found to be highly variable among pastures. We note that many of the biologically small effect sizes that we used for our power analyses are probably smaller than the actual effect sizes we are likely to record in our study, partly because the effect size estimates were sometimes derived from studies that used lower grazing intensities than ours. We might, therefore, find different responses to grazing in our own experiment.

Nonetheless, it is important to acknowledge that we might not be able to detect expected effects. There are two solutions that can be considered. First, this emphasizes the importance of applying a BACI design and sampling the sites before the grazing regime is manipulated (Underwood 1994). This can provide invaluable insight into true differences in change over time among sites (e.g., Schmiegelow et al. 1997).

Second, the hierarchical sampling structure, in which plots are nested within pastures, allows for higher sample sizes at the plot level than at the pasture level (Fig. 1). This can allow for local-scale patterns to be detected even when broad-scale variances are high. Hierarchical study designs should be complemented with an appropriate analytical method that takes advantage of the hierarchical nature of the data, such as mixed-effects models (Fitzmaurice et al. 2004). Importantly, hierarchical designs permit us to capture grazing patterns at large spatial scales, at which herd structure and the location of focal water points are important, without losing resolution to detect gradients of response at finer scales (Stohlgren 2007). This approach also avoids both pseudoreplication and information loss from dropping or averaging data. This is because in hierarchical models, the pastures are treated as the units of replication for independent variables at the pasture scale (e.g., grazing intensity; here, $n = 9$), while plots are treated as the units of replication at the plot scale (e.g., vegetation height; here, $n = 90$), concurrent with controlling for the spatial clustering of plots within pastures.

Researchers need to be willing to concede that it may not be possible to address some questions in which they are interested, within the practical limitations of a study. Efforts should be focused on achievable objectives, with the understanding that response variables that have high intrinsic variation might not respond strongly to management.

The precise estimates of sample sizes that we calculated are sensitive to assumptions such as effect sizes and level of power. Results must be interpreted in the context of this uncertainty. Nonetheless, the tendency for certain variables to require higher sample sizes, and for the continuous design to be more powerful than the categorical design, is relatively insensitive to these assumptions. Our analyses are therefore helpful for comparing relative power of two alternative designs.

Several problems arise from the inflexibility and often inappropriateness of replicated categorical designs for large-spatial scale grazing experiments (Cottingham et al. 2005; Steury and Murray 2005). Spatial constraints often require either reducing the number of replicates or reducing the size of the experimental units to meet requirements of the statistical methods used to analyze categorical designs (Fisher 2000).

These are both compromises that jeopardize the ability to interpret and extrapolate results.

Further, categorical designs result in grazing intensity being treated as a categorical variable, which is biologically inappropriate. Our study area is classified climatically as subhumid shifting towards semiarid under drought conditions (Scott and Suffling 2000). This means that predicted vegetation diversity response to grazing intensity range from a linear to strongly downward concave depending upon the prevailing seasonal climate (Milchunas et al. 1988). Unlike the categorical design, the continuous design will allow us to distinguish among these alternative responses.

IMPLICATIONS

The continuous design, utilized in a modified BACI framework, is superior to the categorical design for evaluating community responses to grazing intensity. We believe that this will typically be the case for large-scale grazing experiments. By spreading grazing treatments across a gradient of disturbance intensities, we may detect both response thresholds as well as gradual shifts in the mixed-grass community.

ACKNOWLEDGMENTS

LITERATURE CITED

Bai, Y., Z. Abouguendia, and R. E. Redmann. 2001. Relationship between plant species diversity and grassland condition. *Journal of Range Management* 54:177–183.

Biondini, M. E., B. D. Patton, and P. E. Nyren. 1998. Grazing intensity and ecosystem processes in a northern mixed-grass prairie, USA. *Ecological Applications* 8:469–479.

Briske, D. D., S. D. Fuhlendorf, and F. E. Smeins. 2003. Vegetation dynamics on rangelands: a critique of the current paradigms. *Journal of Applied Ecology* 40:601–614.

Briske, D. D., S. D. Fuhlendorf, and F. E. Smeins. 2005. State-and-transition models, thresholds, and rangeland health: a synthesis of ecological concepts and perspectives. *Rangeland Ecology and Management* 58:1–10.

Cingolani, A. M., I. Noy Meir, and S. Diaz. 2005. Grazing effects on rangeland diversity: a synthesis of contemporary models. *Ecological Applications* 15:757–773.

Cottingham, K. L., J. T. Lennon, and B. L. Brown. 2005. Knowing when to draw the line: designing more informative ecological experiments. *Frontiers in Ecology and the Environment* 3:145–152.

Dale, B. 1983. Habitat relationships of seven species of passerine birds at Last Mountain Lake, Saskatchewan. Regina, Saskatchewan, Canada: University of Regina. 119 p.

Davis, S. K. 2004. Habitat selection and demography of mixed-grass prairie songbirds in a fragmented landscape [dissertation]. Regina, Saskatchewan, Canada: University of Regina. 129 p.

FAUL, F., E. ERDFELDER, A.-G. LANG, AND A. BUCHNER. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39:175–191.

FISHER, D. S. 2000. Defining the experimental unit in grazing trials. *Journal of Animal Science* 77:1–5.

FITZMAURICE, G. M., N. M. LAIRD, AND J. H. WARE. 2004. Applied longitudinal analysis. Hoboken, NJ, USA: Wiley. 506 p.

FRANK, A. B., D. L. TANAKA, L. HOFMANN, AND R. F. FOLLETT. 1995. Soil carbon and nitrogen of Northern Great Plains grasslands as influenced by long-term grazing. *Journal of Range Management* 48:470–474.

FUHLENDORF, S. D., AND D. M. ENGLE. 2001. Restoring heterogeneity on rangelands: ecosystem management based on evolutionary grazing patterns. *Bioscience* 51:625–632.

FUHLENDORF, S. D., AND F. E. SMEINS. 1997. Long-term vegetation dynamics mediated by herbivores, weather and fire in a *Juniperus–Quercus* savanna. *Journal of Vegetation Science* 8:819–828.

FUHLENDORF, S. D., AND F. E. SMEINS. 1999. Scaling effects of grazing in a semi-arid grassland. *Journal of Vegetation Science* 10:731–738.

GILLEN, R. L., J. A. ECKROAT, AND F. T. MCCOLLUM, III. 2000. Vegetation response to stocking rate in southern mixedgrass prairie. *Journal of Range Management* 53:471–478.

HART, R. H. 2001. Plant biodiversity on shortgrass steppe after 55 years of zero, light, moderate, or heavy cattle grazing. *Plant Ecology* 155:111–118.

HART, R. H., J. BISSIO, M. J. SAMUEL, AND J. W. WAGGONER, JR. 1993. Grazing systems, pasture size, and cattle grazing behavior, distribution, and gains. *Journal of Range Management* 46:81–87.

HART, R. H., M. J. SAMUEL, P. S. TEST, AND M. A. SMITH. 1988. Cattle, vegetation, and economic responses to grazing systems and grazing pressure. *Journal of Range Management* 41:282–286.

HENDERSON, D. C. 2005. The ecology and management of crested wheatgrass invasion [dissertation]. Edmonton, Alberta, Canada: University of Alberta. 137 p.

HILLERISLAMBERS, R., M. RIETKERK, F. VAN DEN BOSCH, H. H. T. PRINS, AND H. DE KROON. 2001. Vegetation pattern formation in semi-arid grazing systems. *Ecology* 82:50–61.

HOWERY, L. D., F. D. PROVENZA, R. E. BANNER, AND C. B. SCOTT. 1998. Social and environmental factors influence cattle distribution. *Applied Animal Behaviour Science* 55:231–245.

JONAS, J., AND A. JOERN. 2007. Grasshopper (Orthoptera: Acrididae) communities respond to fire, bison grazing and weather in North American tallgrass prairie: a long-term study. *Oecologia* 153:699–711.

KOPER, N., AND F. SCHMIEGELOW. 2006. A multi-scaled analysis of avian response to habitat amount and fragmentation in the Canadian dry mixed-grass prairie. *Landscape Ecology* 21:1045–1059.

KOTTEK, M., J. GRIESER, C. BECK, B. RUDOLF, AND F. RUBEL. 2006. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15:259–263.

KREBS, C. J. 1989. Ecological methodology. New York, NY, USA: Harper and Row. 654 p.

LAYCOCK, W. A. 1991. Stable states and thresholds of range condition on North American rangelands: a viewpoint. *Journal of Range Management* 44:427–433.

MICHALSKY, S. J., AND R. A. ELLIS. 1994. Vegetation of Grasslands National Park. Summary Report. D. A. Calgary, Alberta, Canada: Westworth and Associates, Ltd. 111 p.

MILCHUNAS, D. G., W. K. LAUENROTH, AND I. C. BURKE. 1998. Livestock grazing: animal and plant biodiversity of shortgrass steppe and the relationship to ecosystem function. *Oikos* 83:65–74.

MILCHUNAS, D. G., W. K. LAUENROTH, P. L. CHAPMAN, AND M. K. KAZEMPOUR. 1989. Effects of grazing, topography, and precipitation on the structure of a semiarid grassland. *Vegetatio* 80:11–23.

MILCHUNAS, D. G., O. E. SALA, AND W. K. LAUENROTH. 1988. A generalized model of the effects of grazing by large herbivores on grassland community structure. *American Naturalist* 132:87–106.

OSENBERG, C. W., R. J. SCHMITT, S. J. HOLBROOK, K. E. ABU-SABA, AND A. R. FLEGAL. 1994. Detection of environmental impacts: natural variability, effect size, and power analysis. *Ecological Applications* 4:16–30.

QUINN, G. P., AND M. J. KEOUGH. 2002. Experimental design and data analysis for biologists. New York, NY, USA: Cambridge University Press. 520 p.

RIETKERK, M., P. KETNER, J. BURGER, B. HOORENS, AND H. OLFF. 2000. Multiscale soil and vegetation patchiness along a gradient of herbivore impact in a semi-arid grazing system in West Africa. *Plant Ecology* 148:207–224.

ROMO, J. T., AND Y. BAI. 2004. Seed bank and plant community composition, mixed prairie of Saskatchewan. *Rangeland Ecology and Management* 57:300–304.

SASAKI, T., T. OKAYASU, U. JAMSRAN, AND K. TAKEUCHI. 2008. Threshold changes in vegetation along a grazing gradient in Mongolian rangelands. *Journal of Ecology* 96:145–154.

SCHELLENBERG, M. P., N. W. HOLT, AND J. WADDINGTON. 1999. Effects of grazing dates on forage and beef production of mixed prairie rangeland. *Canadian Journal of Animal Science* 79:335–341.

SCHMIEGELOW, F. K. A., C. S. MACHTANS, AND S. J. HANNON. 1997. Are boreal birds resilient to forest fragmentation? An experimental study of short-term community responses. *Ecology* 78:1914–1932.

SCHUMAN, G. E., J. T. MANLEY, R. H. HART, W. A. MANLEY, AND J. D. REEDER. 1999. Impact of grazing management on the carbon and nitrogen balance of a mixed-grass rangeland. *Ecological Applications* 9:65–71.

SCOTT, D., AND R. SUFFLING [EDS.]. 2000. Climate change and Canada's national park system: a screening level assessment. Hull, Quebec, Canada: Environment Canada. 183 p.

SMOLIAK, S., J. F. DORMAAR, AND A. JOHNSTON. 1972. Long-term grazing effects on *Stipa–Bouteloua* prairie soils. *Journal of Range Management* 25:246–250.

STEURY, T. D., AND D. L. MURRAY. 2005. Regression versus ANOVA. *Frontiers in Ecology and the Environment* 3:356–357.

STOHLGREN, T. J. 2007. Measuring plant diversity: lessons from the field. New York, NY, USA: Oxford University Press. 408 p.

STOHLGREN, T. J., M. B. FALKNER, AND L. D. SCHELL. 1995. A modified-Whittaker nested vegetation sampling method. *Vegetatio* 117:113–121.

UNDERWOOD, A. J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications* 4:3–15.