Upland Water and Deferred Rotation Effects on Cattle Use in Riparian and Upland Areas – A Reply to Carter et al. 2017

By Michael R. Guttery and Danny Caudill

On the Ground

- A recent publication by Carter et al. (2017) presents research on the effects of deferred rotation grazing and water provisioning on a suite of environmental variables.
- We detail issues that call into question the validity of the results and conclusions reported by the authors.
- Data were not collected in a scientifically rigorous way.
- Sufficient detail is not presented for the study to be replicated.
- The authors do not adhere to standard statistical definitions or assumptions.
- The study suffers from unaccounted for pseudoreplication.
- The authors draw conclusions beyond the reasonable scope of inference.

Keywords: rotational grazing, water provisioning, statistical sampling, pseudoreplication, reproducible research, scope of inference.

Rangelands 41(2):102–106
doi 10.1016/j.rala.2018.11.005

"To stop injury to the public grazing lands by preventing overgrazing and soil deterioration" (Taylor Grazing Act of 1934). As can be seen in the opening lines of the United States' first and most influential legislation regulating public land grazing, there is a long history of concern about the adverse effects of improperly managed livestock grazing. Undeniably, livestock can contribute to streambank erosion, overgraze rangelands, promote the spread of non-native species, cause shifts in plant community composition, and compete with wildlife for vital resources. Observations of these potential detrimental consequences of ineffective grazing management have inspired generations of research investigating the effects of various grazing management strategies to develop recommendations for the responsible use of natural resources. At first glance, the recent publication by Carter et al.[1] appears to be another contribution toward the goal of effective grazing management, with particular focus on the effects of upland water provisioning and deferred rotation grazing. While these are certainly worthy subjects for research, we found numerous serious issues with the study design and analysis of Carter et al. that call into question the validity of their results and conclusions. In addition, many of the conclusions and recommendations presented by the authors are beyond the scope of their study and are unsupported by data or citations (e.g., a utilization rate of less than 30% is recommended but the study does not evaluate utilization rates). As such, we urge readers to carefully consider the following arguments before accepting or applying the recommendations of Carter et al.

Study Design Issue 1: Purposive Selection of Sampling Sites

With few exceptions, valid statistical sampling involves two aspects: 1) all sites have a known nonzero probability of being selected; and 2) there must be some randomization process involved in selecting sites.[2] Carter et al.[1] states "sampling sites were *chosen* to represent soil map units that covered a majority of the allotment, key range sites identified by BLM, and Utah Division of Wildlife Resources wildlife survey sites" (p. 115; emphasis added). The word *chosen* should shock readers, because if taken at face value this action would clearly be inappropriate because sites should be selected at random so that the investigator cannot unduly influence the outcome of the study. Although the term *chosen* could simply be an unfortunate choice of words, it does not appear to be the case in this instance. The lead author is quoted in a court decision describing site selection on the study area as stating,

> *"It's not like you just pick up the cage, walk over and toss it down somewhere… To me, it's an important decision and I spend a fair amount of time trying to select a site that… represents the true average [of forage production in the area] as best I can determine."*[3]

Clearly it does not appear the data were collected in any rigorously designed fashion, which calls into question the use of inferential statistics on said data to draw broader inference. As is noted by Thompson,[2]

> *"the random or probability selection of samples removes recognized and unrecognized human sources of bias, such as conscious or unconscious tendencies to select units with larger (or smaller) than average values of the variable of interest. Such a [randomization] procedure is especially desirable when survey results are relied upon by persons with conflicting sets of interests – a fish population survey that will be used by fishery managers, commercial fishermen, and environmentalists, for instance. In such cases, it is unlikely that all parties concerned could agree on the purposive selection of a 'representative' sample"* (p. 3).[2]

Study Design Issue 2: Importance of Repeatability

Study replication is the currency of science, thus if a study is not reproducible it is of little worth. There are numerous inconsistencies within Carter et al. regarding the description of the field protocols used in the study which make the study impossible to replicate. For example, the authors state that, "[greenline stubble] Heights were measured on both sides of the stream at approximately 1-m intervals extending 30m up and downstream from the riparian utilization cages" (p. 114).[1] This is the first mention of these riparian utilization cages and we are left to wonder where they came from, although their "selection" does not appear random as per the previous section (similarly, no details are provided regarding the selection of residual plant or bank alteration monitoring sites). Regardless, this sampling description would result in 120 measurements of stubble height at each utilization cage, but the sample sizes reported in table 3 of Carter et al. for stubble height are not multiples of 120.[1] This discrepancy between the number of samples that should exist based on the described methods and the reported sample sizes (table 3 therein) exist for every variable measured in the study (bank alteration = 3 sites × 55 measurement per site × 8 years = 1320 [authors report = 239 + 950 = 1,189]; riparian residuals = 3 sites × 4 frames per site × 8 year = 96 [authors report = 58 + 49 = 107]; upland residuals = 12 sites × 10 frames per site × 8 years = 960 [authors report = 470 + 450 = 920]). Did the authors collect extra samples at some sites, fail to collect samples at some sites, or fail to measure all sites in every year? The authors fail to acknowledge or explain these discrepancies; hence it is impossible to replicate the study.

Analysis Issue 1: Failure to Consider Statistical Assumptions and Definitions

Even if the data collection by Carter et al. was conducted in a rigorous and reproducible manner, there are numerous serious issues with their analyses. Most ecological systems are, admittedly, extraordinarily complex and thereby often confound efforts to apply simplistic experimental or analytical approaches, many of which date to R. A. Fisher's foundational work with agricultural researchers in the early 20th century. Fortunately, the field of statistics has progressed and developed approaches to accommodate common, yet potentially problematic, aspects of many study designs (e.g., repeated measures, nested data, autocorrelation, "nuisance" variables). However, fundamental statistical definitions and assumptions must remain at the forefront of researchers' attention when designing studies and analyzing data. For example, an experimental unit (EU) is generally defined as the smallest unit of experimental space to which a treatment is applied at random.[4,5] Further, the number of EUs in a study directly relates to the total degrees of freedom (i.e., experiment-wide degrees of freedom) in the subsequent analysis, which influences the power of the analysis to detect the presence of a treatment effect if such an effect exists.

Carter et al. do not define what they consider their EU to be, but clearly do not conform, even remotely, to the accepted definition (see Brown and Waller[6] for a similar grazing research critique). From the lacking study description it appears that the "new" water provisioning (p. 114 states several upland water sources existed beforehand) and rotational grazing began in 2010 (p. 114 and table 3 therein), which would make the study a before–after design, and consequently there would be two experiment-wide degrees of freedom. It could be argued that because of longitudinal data collection across multiple years, four replications of each nonrotational (2006–2009) and rotational (2010–2013) grazing exist, leading to eight experiment-wide degrees of freedom. However, the treatments occurred across the same spatial extent as presumably (again unclear study description) all nonrotational grazing, and the rotational grazing treatment was applied throughout the entirety of all pastures that constituted the study area (i.e., a single rotation included the entire study area in a year, with differing ordering of when pastures were grazed across years; e.g., Carter et al. table 1). Hence, the observations of each treatment across subsequent years are not logically independent (they occurred in the same space) and thus likely constitute repeated measures of the same treatment to experimental space. Moreover, most measurements do not occur in all pastures (see fig. 1 therein) as riparian residual and stubble height were only measured in two pastures (2 and 4), and bank alteration measurements only occurred in a single pasture (4). Regardless, from Carter et al. table 3 the authors claim exorbitantly more degrees of freedom than can be determined to exist given the study design and relevant statistical definitions.

Analysis Issue 2: Pseudoreplication

The large disconnect between degrees of freedom and the study design likely exists because the authors have chosen an analysis (t test) that assumes independence of samples and does not acknowledge the hierarchy that clearly exists in the data. In doing so, Carter et al. have committed pseudoreplication, as is described in the seminal paper by Hurlbert.[7] Hurlbert states that, "if 'replicates' are only samples from a single experimental unit, then replicates are not independent. If one uses the data from such experiments to test for treatment effects, then one is committing pseudoreplication." Hurlbert continues on to clarify regarding longitudinal studies, "Successive samples from a single [experimental] unit are so obviously going to be correlated with each other, the potential for spurious treatment effects is very high with such designs." Clearly from the description in Carter et al. there is a hierarchical structure to the data as the study occurred across multiple years (2006–2013; table 1 therein), multiple pastures (fig. 1 therein), multiple measurement sites (fig. 1 therein), along transects at each site (p. 114–115), and at frames along each transect (p. 115).[1] Hence the data become quite complex with longitudinal collection (across years; 2006–2013) and subsampling (frames within transects) of subsampling (transects within sites) within subsampling (sites within pastures) again of subsampling (pastures within treatment) of response variables within the EUs. The true data structure is thus represented by a complex multilevel (i.e., mixed) model structure that has all of the nested subsampling combinations (e.g., transects within sites within pastures within treatment combination levels) crossed with the years, which is clearly exceedingly more complex than can be accommodated by the simplistic t test used by the authors. Failure to account for these various types of dependencies is known to result in artificially precise estimates and consequently exaggerated power of test statistics (i.e., small P values), the bias arising from nonindependent sampling (i.e., pseudoreplication). This egregious misuse of conventional hypothesis testing is a likely example of why P values have been dismissed as inherently incorrect by some (e.g., Krausman[8]), even though the fault is clearly with the practitioner not the test statistic itself (see the American Statistical Association [ASA] statement on the proper use and interpretation of P values[9]).

Generalizability Issue 1: Poor Study Design Resulting from Incomplete Literature Review

We are also concerned that the authors have failed to acknowledge, either intentionally or due to unfamiliarity with relevant research, the body of literature and evidence that runs counter to their conclusions. Moreover, two of the authors have run afoul of this issue before (see Davies et al.[10]). The authors cite the review by Briske et al.[11] to support their unstated hypothesis (again lack of clarity in study description as per previous section) that rotational grazing has little advantage over continuous grazing. However, rotational grazing as a management strategy has more complexity than what was incorporated in the experiments reviewed by Briske et al.[11] (see Briske et al.[13] and Teague et al.[12]). The issue of unconsidered complexity appears to severely limit the utility of the study presented by Carter et al. For example, the authors explicitly state they did not measure changes in plant species (p. 112), whereas a primary goal of rotational grazing is to manage plant-herbivore interactions, or more explicitly, to manage selective grazing in both space and species composition (e.g., Teague et al.,[12] p. 713). Hence the authors' decision to lump all species together clearly ignores the fact that herbivore use of plant resources is complex and not a simple function of total accessible biomass (see Provenza[14,15]). Moreover, realized effects of changes in management regimes typically lag behind implementation because numerous variables are affected simultaneously (e.g., see Teague et al.,[12] p. 709). Consequently, the short duration of Carter et al., particularly post-treatment (4 years), potentially did not allow the system sufficient time to respond to the new management regime. However, even if the authors had appropriately collected data, which they apparently did not, and had appropriately analyzed those data, which they clearly did not, then it would still be vital to acknowledge the observational nature of the study and the context of the results within the larger body of literature. To quote Laplace: "The weight of evidence for an extraordinary claim must be proportioned to its strangeness." The authors attempted to supplant the previous literature and/or evidence, but their data and results are neither extraordinary nor compelling.[19]

Generalizability Issue 2: Scope of Inference

Irrespective of the myriad issues described above, Carter et al. make assertions that are far beyond any reasonable scope of inference their study could support. For example, on the first page of the manuscript the authors state "We did not measure changes in plant species," yet they later conclude: "Sensitive native bunchgrasses are being replaced with increasers and annual forbs." How have the authors arrived at such a conclusion given that they explicitly state that they did not collect the requisite data? The very next sentence in the conclusion goes even further into supposition, stating: "High amounts of erodible bare soil are subject to accelerated erosion, stream channels are incised, and willows are lost." This is the first mention of bare soil (except on page 112 where the authors state that they did not monitor ground cover), incised banks, or willows. The authors have not mentioned measuring any of these variables, so how can they draw such conclusions? The six bulleted recommendations on page 118 of Carter et al. descend into rampant supposition. The first bullet makes recommendations about setting stocking rates, which the authors did not manipulate and which were essentially constant throughout the study (table 2 therein), based on preferred forage species, which the authors explicitly stated they did not measure (p. 112), and current forage consumption rates of livestock, which again the authors did not measure (their cages excluded all grazers including wildlife). The second bullet recommends a 30% utilization rate, yet the authors only observed utilization rates, but did not manipulate them. How can the authors justify this recommendation when this value is nearly half any of the utilization values they observed (see fig. 10 therein)? Moreover, "utilization" by what herbivore species? The authors excluded all grazers with their cages and it seems entirely possible that native grazers alone could account for ≥ 30% of plant biomass use (e.g., see Ranglack et al.[16]). In bullet points three and four we have more of the same. The authors did not manipulate either of the variables they make recommendations on, so how have they arrived at the conclusion that these are the optimal values? In the final two bullet points the authors recommend two strategies (rest and riders) that are different from the strategies that they attempted to evaluate (rotational grazing and upland water provisioning).

Ethical Issue: Acknowledging Potential Conflicts of Interest

We are also quite concerned that the authors seem to have not declared a clear potential conflict of interest. Documents from the aforecited court cases[3,17] clearly show that three of the authors were involved with a legal challenge to grazing practices on the study area. Moreover, a majority of the authors list either Wild Utah Project or Western Watersheds Project as their affiliated institution; those institutions are the appellants in the aforementioned court cases regarding grazing on the study area. This represents a potential conflict of interest. However, we do not intend to imply that a potential conflict of interest should preclude consideration and/or publication of a manuscript, but rather that editors, reviewers, and readers should have been made aware that the authors were involved in relevant litigation so they could judge for themselves the merits of the research and objectivity of the researchers. We also find it quite interesting that validity of the approach and data were a question in the legal decision and we encourage those interested in this data/study to read the court decision as it contains quite relevant information and very pertinent additional context.[3]

Summary

We largely agree with a response, to what appears to be many of these same data and claims, expressed within the previously mentioned court cases, "Appellants' views are nothing more than value-laden opinions" (Western Watersheds Project & Wild Utah Project v. Bureau of Land Management,[17] p. 59). This statement likely extends to the data, as the authors appear to have "chosen" their sampling locations and thus interjected their values and opinions into the data. This is unfortunate because the general topics within Carter et al. are of interest, and evaluating the effects of grazing management strategies is of considerable value. Evaluating different strategies helps to ensure natural resources are managed responsibly. Clearly, no management strategy is ideal for all situations, and all strategies deserve critical evaluation. Nevertheless, management decisions (i.e., deciding which strategy to implement) must be based on rigorous, objective, and reproducible research to the largest extent possible if we expect any kind of predictable (i.e., repeatable) outcome to prevail (see Sells et al.[18]). Science, unlike many endeavors, must bear the burden of evidence, which is established through rigorous study design, data collection, appropriate analyses, objective interpretation, and presentation of reasonable conclusions within the scope of inference. Carter et al. concluded "lack of adequate science-based standards, quantitative monitoring, and enforcement resulted in overuse and degraded conditions." Ironically, the lack of science-based standards (e.g., "chosen" sites), quantitative monitoring (e.g., inappropriate analysis), and enforcement (e.g., thorough peer review) led to the publication of their scientifically flawed manuscript.

In addition to being poor science, we find the publication of Carter et al. to be an example of an unfortunate failure of the peer review process. In summary, Carter et al. suffers from numerous scientifically fatal errors that were missed during the review process:

- The data do not appear to be collected in a manner that conforms to scientific rigor.
- The description of the study does not provide necessary detail such that it could be replicated.
- The analysis is entirely inappropriate given the structure of the study design, which is in itself fundamentally flawed.
- The authors fail to acknowledge literature and evidence that runs counter to their claims.
- The authors make claims and recommendations that are well beyond the scope of inference their data could support even if the data had been collected and analyzed appropriately.
- The authors appear to have failed to acknowledge a clear potential conflict of interest.

Acknowledgments

The authors would like to thank E. Thacker and J. Merickel for conversations that contributed to the development of this paper. We thank C. A. Ribic for helpful comments on an earlier draft of this manuscript. The authors did not receive financial support through grants or contracts to write this paper. Publication fees were generously paid by the Utah Public Lands Policy Coordination Office and the Jack H. Berryman Institute at Utah State University. In preparing this manuscript, the authors declare that they had no known conflicts of interest or motivations other than the desire to ensure that only rigorous science is used to inform natural resource management decisions.

References

1. Carter, J., J.C. Caitlin, N. Hurwitz, A.L. Jones, and J. Ratner. 2017. Upland water and deferred rotation effects on cattle use in riparian and upland areas. *Rangelands* 39:112-118.
2. Thompson, S.K. 1992. Sampling. New York, NY, USA: John Wiley & Sons. 360 p.
3. Bureau of Land Management v. Western Watersheds Project & Wild Utah Project, 191 IBLA 144 (Decided 22 September 2017). Available at: https://www.oha.doi.gov:8080/index.html. Accessed 30 March 2018.

4. Krebs, C.J. 1989. Ecological methodology. New York, NY, USA: Harper Collins. 654 p.
5. Kuehl, R.O. 2000. Design of experiments: statistical principles of research design and analysis. 2nd ed. Pacific Grove, CA, USA: Duxbury Press. 666 p.
6. Brown, M.A., and S.S. Waller. 1986. The impact of experimental design on the application of grazing research results - an exposition. *Journal of Range Management* 39:197-200.
7. Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54:187-211.
8. Krausman, P.R. 2017. P-values and reality. *Journal of Wildlife Management* 81:562-563.
9. Wasserstein, R.L., and N.A. Lazar. 2016. The ASA's statement on p-values: context, process, and purpose. *The American Statistician* 70:129-133.
10. Davies, K., A. Gearhart, M. Vavra, B.W. Schultz, and N. Rimbey. 2016. Longer term rest from grazing: a response to Jones & Carter. *Journal of Rangeland Applications* 3:9-15.
11. Briske, D.D., J.D. Derner, J.R. Brown, S.D. Fuhlendorf, W.R. Teague, K.M. Havstad, R.L. Gillen, A.J. Ash, and W.D. Willms. 2008. Rotational grazing on rangelands: reconciliation of perception and experimental evidence. *Rangeland Ecology and Management* 61:3-17.
12. Teague, R., F. Provenza, U. Kreuter, T. Steffens, and M. Barnes. 2013. Multi-paddock grazing on rangelands: why the perceptual dichotomy between research results and rancher experience? *Journal of Environmental Management* 128:699-717.
13. Briske, D.D., N.F. Sayre, L. Huntsinger, M. Fernandez-Gimenez, B. Budd, and J.D. Derner. 2011. Origin, persistence, and resolution of the rotational grazing debate: integrating human dimensions into rangeland research. *Rangeland Ecology and Management* 64:325-334.
14. Provenza, F.D. 2003. Twenty-five years of paradox in plant-herbivore interactions and "sustainable" grazing management. *Rangelands* 25:4-15.
15. Provenza, F.D. 2003. Foraging behavior: managing to survive in a world of change. Logan, UT, USA: Utah State University. 63 p.
16. Ranglack, D.H., S. Durham, and J.T. du Toit. 2015. Competition on the range: science vs. perception in a bison-cattle conflict in the western USA. *Journal of Applied Ecology* 52:467-474.
17. Western Watersheds Project & Wild Utah Project v. Bureau of Land Management, UT-020-09-01 (Decided 16 May 2013). Available at: https://www.oha.doi.gov:8080/index.html. Accessed 30 March 2018.
18. Sells, S.N., S.B. Bassing, K.J. Barker, S.C. Forshee, A.C. Keever, J.W. Goerz, and M.S. Mitchell. 2018. Increased scientific rigor will improve reliability of research and effectiveness of management. *Journal of Wildlife Management* 82:485-494.
19. Gillispie, C.C., I. Gratton-Guinness, and R. Fox. 1999. Pierre Simon Laplace, A Life in Exact Science. Princetion, NJ: Princeton University Press. 336 p.

Authors are Palmer, AK 99645, USA (micrgutt@gmail.com, Guttery); and Fairbanks, Alaska, USA (Caudill)..