*Commentary*

# Statistical Ritual Versus Knowledge Accrual in Wildlife Science

**FRED S. GUTHERY,**[1] *Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK 74078, USA*

**ABSTRACT** I hypothesized that statistical ritual has supplanted knowledge accrual as the sine qua non of wildlife science. Under the hypothesis, I deduced occurrence of 1) significance testing of the obvious and inconsequential, 2) quantitative debasement of research problems, and 3) publication of papers that largely lacked information but were methodologically impeccable. Articles in past and recent wildlife literature fit the deductions and supported the hypothesis. Thus, wildlife science is operating inefficiently because quantitative formalities are supplanting ecological information in technical articles. This problem can be corrected by a change of mindset in authors, referees, and editors. The change entails less emphasis on quantitative ritual and more emphasis on information that aids in understanding and explaining nature and managing wildlife. (JOURNAL OF WILDLIFE MANAGEMENT 72(8):1872–1875; 2008)

DOI: 10.2193/2008-179

**KEY WORDS** Akaike-best model, cargo cult science, model selection, pseudodifference, quantitative ritual, significance testing.

"*We know that Thoreau . . . feared that man was becoming the tool of his tools, which can, alas, include ideas*" *(Eiseley 1969:142).*

"In the South Seas there is a cargo cult …. During [World War II] they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for [the controller] to sit in, . . . and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land." (Feynman 1985:340).

From the above description, Feynman (1985:340) coined the phrase "cargo cult science" that follows "all the apparent precepts and forms of scientific investigation," but that misses something essential. His description of cargo cults metaphorically brings to light a property of the human condition that may be inimical in science. That property is ritualism (runways, fires, huts) as a means of obtaining some desideratum (cargo) in a manner that may seem ludicrous to an outside observer. Such behavior and the tribalism it engenders and supports probably are products of human evolution (Wilson 2004). Scientists, as humans, may be genetically predisposed to ritualism and orthodoxy (Brothers 1997, Guthery 2008), although by dint of training and tradition they try to bring logic and objectivity to the fore and to keep their social urges subdued. However, these urges are not inevitably subjugated by logic and objectivity. Widespread reporting of naked *P*-values (Anderson et al. 2000), which have little or no heuristic value, provides an example of ritualistic behavior trumping knowledge accrual in wildlife science (Guthery 2008).

What I call quantitative ritual many scientists would call good statistical practice. Yet when application becomes more or less involuntary, good practice takes on properties of ritual in a religious sense. Then adherence to procedure seemingly becomes more important than reasoning and information acquisition in the conduct of science. I hasten to point out that quantitative practice is mostly, but not completely, an indifferent set of algorithms and that human application of the algorithms is the problem.

Herein I evaluate the hypothesis that ritualistic application of quantitative methods (significance testing and model selection) has replaced knowledge accrual as the sine qua non of wildlife science. If the hypothesis holds, then our literature will contain verbal matter that may be viewed as manifestations of ritualism because it provides little or no knowledge or impedes knowledge accrual. Examples include 1) statistical significance testing of the obvious and reports of significant findings on the inconsequential, 2) quantitative debasement of research problems, and 3) publication of results with little or no informational substance (or even outright error) that are methodologically de rigueur.

## SIGNIFICANCE TESTING

Johnson (1999) noted inane applications of null hypothesis significance testing (NHST), such as tests to determine whether logging affects density of trees in a forest. Apparently mindless application of NHST continues with surprising frequency. Recently, I have seen significance tests to determine whether irrigation treatments affect seed germination (Cornaglia et al. 2005), body water affects mass of a carnivore (Hwang et al. 2005), mowing affects height of herbaceous vegetation (Washburn and Seamans 2007), and clear-cutting affects density of trees in a forest (Morris and Maret 2007). One paper (unpublished) reported that roosting time of a bird species was correlated ($P < 0.001$) with sunset time. I attended a seminar where the speaker reported sickness reduces the number of days people work ($P < 0.05$). That *P*-value sure was comforting from a scientific standpoint. There is no point other than adhering to ritual in doing statistical tests of the inevitable. By applying such tests, wildlife science parodies itself.

A problem related to statistical tests of the inevitable is the inevitability of finding statistical significance. I define a pseudodifference as a significant difference that is opera-

[1] *E-mail: fred.guthery@okstate.edu*

tionally and biologically inconsequential or even trivial. Large sample sizes are appropriately extolled in wildlife science, but such samples have the insidious property of generating pseudodifferences ($P < 0.05$). As sample size grows, any difference(s) between ≥2 means, no matter how small the difference(s), ordinarily will become statistically significant (Johnson 1999). Put differently, all nonzero differences between any 2 means are statistically significant in the limit. In a related vein, I recently did a logistic regression analysis (unpublished) on properties of used versus random points. The statistical-software-generated estimate of one coefficient was 0.000 ($P < 0.001$, $n = 751$). This could be termed, somewhat paradoxically, a statistically significant null effect.

Wildlife scientists rarely recognize and acknowledge pseudodifferences. Recognition would entail statements in Discussion such as, "Although $x$ differed from $y$ ($P < 0.001$), the significance probably reflected sampling intensity rather than any meaningful property of the study population; moreover, the difference was so small as to be ecologically inconsequential." Ritual precludes such honesty. Asymmetrically, wildlife scientists often rue lack of significance on the basis of sample size. This asymmetry implies a quest for statistical significance, which seems ritualistic because the purpose of science is to gather knowledge, not to seek after artificial constructs such as small *P*-values or Akaike-best models.

The continued existence per se of NHST and *P*-values is evidence of their ritualization because these constructs have long been criticized as quasi-scientific (Morrison and Henkel 1970). "Null hypotheses of no difference," for example, "are usually known to be false before the data are collected; when they are, their rejection or acceptance simply reflects the size of the sample and the power of the test, and is not a contribution to science" (Savage 1957:332–333; need I point out that this quote from a statistics journal is >50 yr old?). "The usefulness of *P*-values is quite limited," argued Anderson and Burnham (2002:915), "and we continue to suggest that these procedures be euthanized." Yet *P*-values continue to be used heavily by authors and mandated by referees and editors. Science or ritual? Authors (e.g., Lloyd and Slater 2007) are beginning to fuse significance testing and model selection into a weird, somewhat self-contradictory amalgam that smacks of paradox. Does significance testing trump model selection or vice versa?

## QUANTITATIVE DEBASEMENT OF RESEARCH PROBLEMS

The advent of model selection using information-theoretic approaches and the Akaike Information Criterion (AIC; Burnham and Anderson 1998) brought with it opportunity to quantitatively debase research problems. By quantitative debasement, I mean adding unnecessary and perhaps misleading quantitative manipulations. The problem ranges from imposing model selection on simple problems easily dealt with using descriptive statistics to unbridled formulation of models ranging from simple to hopelessly complex (Guthery et al. 2001, 2005). Model selection lends itself to hopelessly complex models, for given $n$ independent variables, there are $2^n - 1$ models that contain a range of $1 - n$ of the variables. Variable transformations (e.g., squaring or taking logarithms) and interaction effects provide an additional lode of variables. Model selection, as applied, can often be described as formally structured guesswork.

As a result, an Akaike-best model often represents the most plausible incomprehensible model because it cannot be interpreted. An example is a best-regression model for predicting habitat selection (Whitaker et al. 2006); the model contained 11 parameters, including coefficients for 1 log-transformed variable, 3 2-way interactions, and 1 3-way interaction. Human beings cannot understand this model. Try to imagine pairing a 2-way interaction with a 3-way interaction and graphically predicting the response variable while holding 9 other variables constant at some arbitrary value. Imagining this is impossible, and so is graphing it. This same paper reported an Akaike-best model with 24 parameters that explained 5% of the variation in the data. Thus, the model was simultaneously incomprehensible and worthless. I conjecture noninformation of this type appears in print at least in part because it is based on impeccable quantitative ritual. If this conjecture holds, it implies that authors, referees, and editors often judge papers based more on adherence to quantitative protocol than on contributions to knowledge. I recognize, of course, that authors also see adherence to protocol as strengthening a contribution to science.

As an aside, authors almost invariably report the variables appearing in the Akaike-best model. This is as it should be. Some authors stop at that point, which renders a paper largely devoid of information. Knowing a variable appeared in a best model is no more helpful in understanding ecological process than knowing the French executed Mata Hari in 1917. Authors can contribute to knowledge by providing interpretations of how a variable influences ecological process. For example, regression coefficients have biological meaning as well as statistical interpretations (Guthery and Bingham 2007).

## RESULTS LACKING INFORMATION

Quantitative ritualism can obscure and even bastardize information contained in data. Consider a multiple regression exercise that involved a large set of independent variables (Farrell and Tappe 2007). Some of the independent variables were straightforward measurements and others were fairly abstruse (e.g., a patch-shape index). The abstruse variables were the first sign that interpretable information had been compromised. Each independent variable was tested for normality and transformed as needed to fit an assumption of multivariate normality. Transformations included $\ln(x)$, $\arcsin(x)$, $\arctan(x)$, $x^{1/2}$, $\arcsin(x^{1/2})$, and $\arctan[\ln(x)]$. These data transformations are tantamount to data mutilations because they render the variables largely mysterious to people with normal intelligence. For example, $\arctan[\ln(x)]$ is the angle with a tangent equal to the natural logarithm of *x*. How does one interpret this transformation ecologically? What if one finds arctan[ln(*x*)] is a good predictor of the dependent variable? I think it is impossible to explain how or why the transformed variable is a good predictor.

Good statistical practice calls for reducing the number of independent variables and eliminating correlated variables in developing a regression model (Kleinbaum and Kupper 1978, Afifi and Clark 1984); Farrell and Tappe (2007) proceeded accordingly. Principal component analysis is a method of accomplishing these goals and it is possible to generate factor scores that are not correlated (satisfying the assumption of independence) and to model the dependent variable on these multivariate scores. The trouble is, the resulting regression model is beyond human comprehension except possibly in the case of extremely simple and orderly data. How can one possibly give a biological interpretation of a regression coefficient in a multivariate model (Guthery and Bingham 2007) derived from factors scores when the scores themselves are products of multivariate modeling? We have a multivariate model of a multivariate model. Interpretation is impossible, even for geniuses. Although it does not necessarily follow that the gyrations described above are ritualistic simply because the results are uninformative, ritualism remains a viable explanation.

## DISCUSSION

I tested the hypothesis that quantitative ritual has supplanted knowledge accrual as the sine qua non of wildlife science. Under the hypothesis, I predicted the appearance in our literature of mindless significance testing, quantitative debasement of research problems, and papers lacking information that were quantitatively impeccable. I provided examples of each prediction, which supports the hypothesis under hypothetico–deductive (H–D) logic. This outcome is consistent with the general tendency of human beings to transmute means (e.g., quantitative analysis) and ends (e.g., knowledge accrual).

In the real world, however, H–D logic is flawed because different hypotheses may lead to the same deductively derived conclusion (Guthery 2008). What else might explain the examples I provided? One possibility is incompetence on my part. Perhaps I should not accept without a supporting significance test that mowing reduces height of vegetation. Perhaps most wildlife scientists can understand an 11-parameter model containing a 3-way interaction, but I truly doubt any can.

We have gone too far in following "all the apparent precepts and forms of scientific investigation" (Feynman 1985:340) and in doing so we have placed quantitative corroborees above knowledge accrual as the priority of our science. Wildlife science is migrating into quantitative pomp and circumstance, which negates its primary goals of describing, explaining, and understanding nature. This leads to the problems I discussed above. I do not know the breadth and depth of these problems, but I suspect they are common and their frequency and intensity are increasing.

To forestall further quantitative degradation of our science, practitioners will have to recognize the problem and take corrective actions. Some actions such as eliminating "needlessly complicated and confusing statistical methodology" (Murtaugh 2007:56) are mechanical though perhaps contrary to prevailing tribalism. Many scientists see rigor as synonymous with statistical complexity and simple message as synonymous with laxness. These scientists have a somewhat religious outlook on the practice of wildlife science, and what its practitioners "should" do. Few transcendental mandates ("shoulds") cover all of wildlife science and none involves quantitative analyses (Guthery 2008).

I sense an upwelling of dissatisfaction with the direction wildlife science has turned. The sense arises from discussions with colleagues and observations of the literature. "Lust for statistics" is a symptom that wildlife management has broken free of its natural-history roots (Herman 2002:933). Though rare, papers lacking significance testing and model selection recently have returned to the wildlife literature (Peterson 2001, Guthery et al. 2005, Rader et al. 2007). The work of Aldo Leopold, Paul L. Errington, Herbert L. Stoddard, Margaret Morse Nice, Charles Elton, Francis Hamerstrom, Charles Darwin, and many others shows that obeisance to statistical testing or model selection is not requisite for accruing knowledge about nature.

### Management Implications

If quantitative ritual has replaced knowledge accrual as the essential condition of wildlife science, then our science is operating inefficiently as a mechanism for knowledge accrual. The ratio of information to noninformation in recent technical articles is often low, possibly declining, and in some cases approaching zero. Correcting the problem depends primarily on a change in the mindsets of authors, referees, and editors. They must come to see interpretable ecological information that improves our understanding of nature, not quantitative ritual, as the goal of research and publishing. Given this change, the messages contained in ecological data will be more transparent and more readily applicable by the community of wildlife scientists and managers.

### Acknowledgments

S. W. Dunkin, J. Shaw, L. A. Brennan, J. J. Lusk, M. K. Owens, and M. J. Peterson critiqued a draft of this manuscript. The Bollenbach Endowment, Game Bird Research Fund, Department of Natural Resource Ecology and Management, and Oklahoma Agricultural Experiment Station provided financial support. The Oklahoma Agricultural Experiment Station approved this paper for publication.

## LITERATURE CITED

Afifi, A. A., and V. Clark. 1984. Computer-aided multivariate analysis. Lifetime Learning, Belmont, California, USA.

Anderson, D. R., and K. P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. Journal of Wildlife Management 66: 912–918.

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null

hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912–923.
Brothers, L. 1997. Friday's footprint: how society shapes the human mind. Oxford University Press, New York, New York, USA.
Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference. Springer, New York, New York, USA.
Cornaglia, P. S., G. E. Schrauf, M. Nardi, and V. A. Deregibus. 2005. Emergence of dallisgrass as affected by soil water availability. Rangeland Ecology & Management 58:35–40.
Eiseley, L. 1969. The unexpected universe. Harcourt Brace Javonovich, New York, New York, USA.
Farrell, M. C., and P. A. Tappe. 2007. County-level factors contributing to deer–vehicle collisions in Arkansas. Journal of Wildlife Management 71: 2727–2731.
Feynman, R. P. 1985. Surely you're joking, Mr. Feynman. W. W. Norton, New York, New York, USA.
Guthery, F. S. 2008. A primer on natural resource science. Texas A&M University Press, College Station, USA.
Guthery, F. S., and R. L. Bingham. 2007. A primer on interpreting regression models. Journal of Wildlife Management 71:684–692.
Guthery, F. S., L. A. Brennan, M. J. Peterson, and J. J. Lusk. 2005. Information theory in wildlife science: critique and viewpoint. Journal of Wildlife Management 69:457–465.
Guthery, F. S., J. J. Lusk, and M. J. Peterson. 2001. The fall of the null hypothesis: liabilities and opportunities. Journal of Wildlife Management 65:379–384.
Guthery, F. S., A. R. Rybak, S. D. Fuhlendorf, T. L. Hiller, S. G. Smith, W. H. Puckett, Jr., and R. A. Baker. 2005. Aspects of the thermal ecology of bobwhites in north Texas. Wildlife Monographs 159.
Herman, S. G. 2002. Wildlife biology and natural history: time for a reunion. Journal of Wildlife Management 66:933–946.
Hwang, Y. T., S. Larivière, and F. Messier. 2005. Evaluating body condition of striped skunks using non-invasive morphometric indices and bioelectrical impedance. Wildlife Society Bulletin 33:195–203.
Johnson, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763–772.
Kleinbaum, D. G., and L. L. Kupper. 1978. Applied regression analysis and other multivariable methods. Duxbury Press, North Scituate, Massachusetts, USA.
Lloyd, J. D., and G. L. Slater. 2007. Environmental factors affecting productivity of brown-headed nuthatches. Journal of Wildlife Management 71:1968–1978.
Morris, K. M., and T. J. Maret. 2007. Effects of timber management on pond-breeding salamanders. Journal of Wildlife Management 71:1034–1041.
Morrison, D. E., and R. E. Henkel, editors. 1970. The significance test controversy. Aldine, Chicago, Illinois, USA.
Murtaugh, P. A. 2007. Simplicity and complexity in ecological data analysis. Ecology 88:56–62.
Peterson, M. J. 2001. Northern bobwhite and scaled quail abundance and hunting regulation: a Texas example. Journal of Wildlife Management 64:828–837.
Rader, M. J., T. W. Teinert, L. A. Brennan, F. Hernàndez, N. J. Silvy, and X. Ben Wu. 2007. Identifying predators and nest fates of bobwhites in southern Texas. Journal of Wildlife Management 71:1626–1630.
Savage, I. R. 1957. Nonparametric statistics. Journal of the American Statistical Association 52:331–344.
Washburn, B. E., and T. W. Seamans. 2007. Wildlife responses to vegetation height management in cool-season grasslands. Rangeland Ecology & Management 60:319–323.
Whitaker, D. M., D. F. Stauffer, G. W. Norman, P. K. Devers, T. J. Allen, S. Bittner, D. Buehler, J. Edwards, S. Friedhoff, W. M. Giuliano, C. A. Harper, and B. Tefft. 2006. Factors affecting habitat use by Appalachian ruffed grouse. Journal of Wildlife Management 70:460–471.
Wilson, E. O. 2004. On human nature. Harvard University Press, Cambridge, Massachusetts, USA.

*Associate Editor: Kroll.*