Studying Biodiversity: Is a New Paradigm Really Needed?

JAMES D. NICHOLS, EVAN G. COOCH, JONATHAN M. NICHOLS, AND JOHN R. SAUER

Authors in this journal have recommended a new approach to the conduct of biodiversity science. This data-driven approach requires the organization of large amounts of ecological data, analysis of these data to discover complex patterns, and subsequent development of hypotheses corresponding to detected patterns. This proposed new approach has been contrasted with more-traditional knowledge-based approaches in which investigators deduce consequences of competing hypotheses to be confronted with actual data, providing a basis for discriminating among the hypotheses. We note that one approach is directed at hypothesis generation, whereas the other is also focused on discriminating among competing hypotheses. Here, we argue for the importance of using existing knowledge to the separate issues of (a) hypothesis selection and generation and (b) hypothesis discrimination and testing. In times of limited conservation funding, the relative efficiency of different approaches to learning should be an important consideration in decisions about how to study biodiversity.

Keywords: biodiversity, deduction, induction, machine learning, science

There is much yet to learn about the structure and dynamics of ecological systems, and the conservation of biodiversity depends on such knowledge. Given this challenge, what is the recommended approach to such learning? Kelling and colleagues (2009; also see Caruana et al. 2006, Hochachka et al. 2007) answered this question by proposing a new paradigm for biodiversity studies. The basic outline of their recommendation is as follows: (a) Organize large amounts of data from multiple sources. (b) Analyze these data using automated, computer-based approaches designed to discover complex patterns. (c) Develop hypotheses about the processes that generated these patterns. Kelling and colleagues (2009) referred to this approach as *data intensive science* and claimed that it “allows scientists to analyze bigger and more complex systems efficiently” (p. 613).

Kelling and colleagues (2009) contrasted this new (but see Tukey 1962, 1977) paradigm with “more traditional approaches of hypothesis generation and experimental testing” (p. 613). These traditional approaches begin with an investigator armed with one or more hypotheses about how a system works. Models associated with these hypotheses are developed in order to make predictions. The investigator then focuses on collecting data that can be compared with predictions in order to test the single hypothesis or to discriminate among competing hypotheses. The resulting confrontation between data and model-based predictions leads to more or less confidence in associated hypotheses, depending on the correspondence of predictions and observations (Hilborn and Mangel 1997, Williams et al. 2002). Kelling and colleagues (2009) referred to this approach as “inherently limiting” (p. 613) and “argue[d] that a data-driven approach is necessary in biodiversity studies because of the complexity of ecological systems, particularly when viewed at large spatial and temporal scales” (p. 618).

The above approaches contrasted by Kelling and colleagues (2009) led to an asymmetric, apples-to-oranges comparison. Their data-driven approach is explicitly focused on hypothesis generation and selection (i.e., what hypotheses should be considered). Inferences about the degree of correspondence between their new hypotheses and natural processes would still be based on traditional approaches to hypothesis testing (Kelling et al. 2009). The traditional, knowledge-driven approach to science also includes the development of new hypotheses, but this development is based primarily on existing knowledge. This knowledge-based approach is then conditional on new or existing hypotheses and is focused on testing or discriminating among them.

We begin with a discussion of how scientists generate new hypotheses and then select which hypotheses they choose to test, because we view these issues as the real focus of Kelling and colleagues (2009). We believe that data-driven and knowledge-driven approaches to hypothesis generation and selection are best viewed as complementary, in the sense that both approaches have the potential to produce useful hypotheses. However, finite time, effort, and resources require the investigator to determine the appropriate allocation of these resources. An optimal (or even informed) allocation of time and effort to these two approaches to hypothesis generation requires information about the costs and benefits of each approach. The parameters associated with likely

BioScience 62: 497–502. ISSN 0006-3568, electronic ISSN 1525-3244. doi:10.1525/bio.2012.62.5.11

benefits (e.g., probabilities of ultimate hypothesis utility) are characterized by such great uncertainty that we make no attempt to provide such guidelines. Instead, we follow the basic advocacy approach of Kelling and colleagues (2009), posing the question in our title and providing our opinions about whether the new paradigm is really more likely than traditional approaches to be useful in generating hypotheses about how ecological systems work. Following this discussion of hypothesis generation and selection, we return to the topic of hypothesis testing and discrimination and consider some of the arguments posed by Kelling and colleagues (2009) about the limited utility of these knowledge-based approaches for the study of complex biological systems.

Hypothesis generation and selection

Kelling and colleagues (2009) advocated machine-learning methods with large amounts of data to discover complex patterns from which hypotheses could be developed. They argued the necessity of this approach for biodiversity studies, as opposed to "a more traditional knowledge-driven approach" (p. 618). Our view of a knowledge-based approach is one in which hypotheses selected for testing (a) come directly from existing scientific literature, (b) are deduced from hypotheses and theory in that literature, or (c) are deduced from whatever first principles or knowledge the investigator may have that is relevant to the focal system. The second and third sources involve the generation of new hypotheses.

Existing hypotheses. A stated goal of data-driven science is "the discovery of complex patterns" (p. 613), which can then be used as bases for hypothesis generation (Kelling et al. 2009). A tenet of knowledge-driven science is that much can be learned by testing preexisting hypotheses. We remind the reader of the decades-old admonishment of Romesburg (1981) that investigations in wildlife science were being dominated by retrospective approaches in which investigators collected data and then developed stories about the processes that generated them. Such work was thus focused on hypothesis generation. Romesburg (1981) argued that hypotheses were not then in short supply and that scientists should be expending their efforts on the more useful task of discriminating among existing plausible hypotheses. Indeed, if we look at the ecological literature dealing with macroecology and species-distribution patterns (a focus of Kelling et al. 2009), we find books and reviews filled with interesting hypotheses that merit our efforts to test and discriminate (e.g., Brown 1995, Brown et al. 1996, Gaston and Blackburn 2000, Gaston 2003). Of course this does not mean that no important patterns remain to be discovered, and we argue below that such discovery is more likely to result from knowledge-based approaches focused on important scientific questions or conservation problems.

New hypotheses. For cases in which existing hypotheses are deemed inadequate (i.e., are not consistent with reality, as is judged from relevant tests and observations), the investigator must consider approaches to generate new hypotheses. Both the new paradigm and the traditional approaches described by Kelling and colleagues (2009) have very old conceptual bases in the processes of induction and deduction, respectively. In his essay published in 1919 in the German newspaper *Berliner Tageblatt* and devoted to the relative merits of inductive and deductive approaches, Einstein (translated in Adam 2000) described the basic step of the inductive method as follows: "Singular facts are chosen and grouped in such a way that the law of nature, which connects them, becomes evident" (pp. 34–35) However, Einstein then went on to indicate that this approach had not proven useful in scientific investigations: "The really great progress of natural science arose in a way [that] is almost diametrically opposed to induction" (Adam 2000, p. 35). We share this view and will elaborate on this opinion.

With respect to the three basic steps of the new paradigm outlined by Kelling and colleagues (2009), we recognize the abilities to organize large amounts of data, to use computer algorithms to discover patterns in them, and even to then develop hypotheses about the processes that generated those patterns. However, there are potential difficulties with conducting these last two steps in ways that are not driven by existing knowledge. The ability to identify patterns, for example, requires some prior knowledge of what constitutes the absence of a pattern. The final step of the new paradigm—developing useful hypotheses about the processes that generate patterns—will virtually always be based on prior knowledge. In addition, although ecologists are very good at the generation of hypotheses, it can be very difficult to decide which hypotheses are worthy of strong consideration. This inherent problem can be characterized by a statement that has become known as *Phaedrus's law*: "The number of rational hypotheses that can explain any given phenomenon is infinite" (Pirsig 1974, p. 107). If we are able to develop strong hypotheses (in the sense that they are likely to provide good approximations of reality) to explain a pattern, it will typically be because we have prior knowledge of or hypotheses about the studied system, not because the pattern itself somehow leads to viable hypotheses about its generating processes. But if we had such prior knowledge and hypotheses about the system in the first place, we would have used these to guide our search for patterns. As was noted by Einstein, "If, in fact, the researcher approaches things without any preconceived opinion, how could he even pick out facts from the immense abundance of complicated experience [that] are simple enough that the laws become apparent?" (Adam 2000, p. 35).

Proponents of the new paradigm claim that this approach is necessary for studying complex systems with large numbers of variables and drivers (Kelling et al. 2009). However, it is not obvious at all to us why approaches based on pattern recognition should be more useful with systems that are especially large and complex. If anything, we would think that hypotheses and prior knowledge are even more important when complex systems are studied. Certainly, increased size and complexity translate to greatly increased numbers of potential patterns and require more diligence to separate out patterns that are either spurious (Cole 1957, Anderson et al. 2001) or of little relevance to important ecological questions and conservation problems (Hand 2005). At a minimum, the burden is on proponents of data-based pattern recognition to demonstrate its superiority in the face of large size and great complexity. Faced with large numbers of patterns and hypotheses, it would seem that we should use any sort of prior knowledge that might help us sift through them and discard those that are unlikely to be plausible or useful.

The dynamics of nonlinear systems represent one area in which the difficulties in distinguishing random noise from the output of complex models are well known. Signals generated by purely stochastic processes can appear identical to those generated by a complex nonlinear system when they are viewed through the lens of certain data-analysis tools. Sugihara and May (1990), for example, noted that a tailored stochastic process can mimic the patterns found in a given complex deterministic process. Indeed, a number of authors have since advocated careful hypothesis testing (the classical hypothetico-deductive approach) for distinguishing between these two potential explanations of the data (e.g., Stone 1992). One wonders how much time has been devoted to looking for plausible hypotheses to explain the patterns in purely stochastic signals.

Hypothesis selection. Scientists therefore have the flexibility to select existing hypotheses to test or to develop new hypotheses. What sorts of considerations should be important to this decision? Science has long been viewed as a progressive endeavor (Fara 2009), a perspective that seems very sensible. A translation from Descartes (Lafleur 1960, p. 46) reads, "I hoped that each one would publish whatever he had learned, so that later investigations could begin where the earlier had left off." Indeed, it simply does not seem reasonable to try to conduct ecological investigations in a manner that explicitly ignores more than a century of ecological science. "Wish I didn't know now what I didn't know then" (Seger 1980) does not seem to be an appropriate attitude for a scientific investigator. The standard way to bring prior knowledge to bear on any problem is to develop hypotheses based on such knowledge, an approach that is consistent with Kuhn's (1962) "normal science." We do not deny the possibility of important patterns and hypotheses that do not emerge from prior knowledge, but we do not view methods that focus on their discovery as an efficient approach to increase existing knowledge.

A referee of this manuscript asked why an existing hypothesis that has been subjected to efforts to falsify it should merit more attention than a newly generated hypothesis. Consider two similar-sized groups of new (untested) hypotheses that have similar a priori chances of including a few useful hypotheses (i.e., they provide good approximations of reality). Let the members of one group be subjected to reasonable tests that result in discarding some hypotheses because they are clearly inconsistent with data (and reality). Using Popper's (1972) "natural selection of hypotheses" analogy, the latter group has undergone a process akin to mortality selection, with the result that remaining members of this group exhibit a higher average fitness than do the members of the initial group that has undergone no such selection. A randomly selected hypothesis from the tested group thus has a greater probability of being useful than a randomly selected hypothesis from the untested, new group. Most scientific disciplines recognize that existing, tested hypotheses have added value and merit documentation, and scientific books and journals bear testament to this view.

Kelling and colleagues (2009) expressed the hope that their new methods would be useful for dealing with threats to biodiversity and developing science-based environmental policies. We believe that an interest in conservation should provide a means of further narrowing the focus of scientific investigations. For conservation, we are not just interested in the large set of newly generated or otherwise existing hypotheses about variation in species distribution patterns. Instead, we will probably focus on hypotheses about the sources of variation in species distribution patterns that are associated with human activities or that are potentially influenced by management actions. Indeed, the existence of prior hypotheses permits a formal valuation of different kinds of information for decisionmaking (e.g., Raiffa and Schlaifer 1961, Williams et al. 2002, Runge et al. 2011), further sharpening our focus and increasing the efficiency of our data-gathering efforts. Our prior hypotheses are thus based on the explicit needs of conservation programs and provide an effective means of insuring appropriate focus of our science—in this case, on species conservation.

Hypothesis testing and discrimination

Kelling and colleagues (2009) contrasted their new paradigm with the more traditional "expert-centered" approach of hypothesis generation and experimental testing. They criticized the traditional approach as inadequate for dealing with complex problems of large scale. True manipulative experimentation requires the random assignment of treatments to replicate experimental units (Fisher 1947) and is widely recognized as leading to the strongest possible inferences. However, hypothetico-deductive approaches to science do not require experimentation. Frequently, true manipulative experiments cannot be conducted in ecology and conservation biology, for a variety of reasons. Manipulation, in the absence of randomization or replication or both, yields constrained designs that can be very useful, despite their yielding weaker inferences than true experiments (see, e.g., Green 1979, Skalski and Robson 1992, Nichols 2001, Williams et al. 2002). Observational studies, which lack any manipulations, can still yield relevant inferences, although there are two distinct approaches to inference with observational data—one primarily deductive and one primarily inductive. The deductive approach has been labeled *prospective* and requires prior specification of hypotheses that can be evaluated with the observational data (Nichols 2001, Yoccoz et al. 2001, Williams et al. 2002). The inductive approach has been labeled *retrospective* (Romesburg 1981, Nichols 2001, Williams et al. 2002) and requires an examination of the observational data and subsequent development of hypotheses about the processes that generated them. We believe that the strength of inference for these approaches is typically governed by the following hierarchy: Manipulative experiments provide stronger inferences than constrained design manipulation, which provides stronger inferences than prospective observation, which in turn provides stronger inferences than retrospective observation (also see Nichols 2001, Yoccoz et al. 2001, Williams et al. 2002). The first three approaches are consistent with knowledge-driven science and can lead to changes in the relative degrees of faith that we have in the prior hypotheses that were selected for investigation. The latter inductive approach can be useful for hypothesis generation, but we would be foolish to place much faith in the resulting new hypotheses because of the reality of Phaedrus' law.

The above inequality in strength of inference applies even when data are well suited to a retrospective observational approach. However, this approach to inference, unlike the others listed above, is characterized by a limited, at best, ability to tailor data collection to the intended data use. Indeed, Kelling and colleagues (2009) emphasized the ability of their data-driven approach to utilize observational data collected in diverse ways for diverse purposes. The existence of prior hypotheses in traditional, knowledge-driven approaches permits substantial increases in the strength of inference from study design. Design ensures that data-collection activities are tailored to the primary uses to which the data will be put. In the case of large-scale ecological processes, two important sources of variation in raw data, and hence two foci of design, are geographic sampling and detection (e.g., Lancia et al. 1994, Yoccoz et al. 2001, Thompson 2002, Williams et al. 2002). Because data can seldom be collected at all locations about which an inference is to be made, surveyed locations must be selected in ways that permit inferences about locations not surveyed and that permit inferences about the hypotheses of interest. *Detection* refers to the feature of most ecological sampling that all individuals of any species of interest that are present at a location are seldom detected in survey efforts at that location. Some fraction of individuals is nearly always missed, and many study designs include the collection of extra data to estimate this fraction or to otherwise permit inferences in the presence of nondetection. Raw data from ecological sampling programs are therefore a reflection of two processes—the design and sampling processes that include the determination of the surveyed locations and the method or methods of data collection—and true ecological processes that reflect actual variation in the distributions of individuals and species over space and time. Observational approaches that lack an appropriate design yield data that confound sampling and ecological processes, such that perceived patterns may be produced by one or both types of processes. For example, Hosseini and colleagues (2004) described a demographic and behavioral model to explain strongly seasonal dynamics in the apparent prevalence of mycoplasma infections in house finches (*Carpodacus mexicanus*) (Altizer et al. 2004). However, Faustino and colleagues (2004), Jennelle and colleagues (2007), and Conn and Cooch (2009) each found that the detection probability of house finches varied with the birds' disease state; in addition, there was a strong seasonality in the state-specific detection probabilities, which created uncertainty about the degree to which the apparent pattern reflected interesting biology or sampling uncertainty.

Even when survey design is not tailored to specific questions, it is very relevant to data analysis under a hypothetico-deductive approach. Statistical analyses must accommodate design issues (data-collection locations and methods) in order to yield inferences that are appropriate to the questions and quantities of interest. An absence of study design and a failure to accommodate design issues in data analysis create a problem, in that resulting raw data and the associated analytic results now represent complicated functions of sampling and ecological processes. For example, it will be difficult to determine whether a detected pattern is worthy of investigation, since it will reflect variation in surveyed locations and detection probabilities, as well as possible variation in animal numbers or occupancy. Indeed, it has been argued that patterns detected in data resulting from some volunteer-based monitoring programs reflect a mapping of survey effort rather than of quantities of ecological interest. Therefore, study design represents a strong advantage of studies focused on prior hypotheses. Even in the absence of tailored designs, issues associated with the locations and methods of data collection must be accommodated in analyses focused on ecological hypotheses and patterns. Once again, the identification of prior hypotheses and the associated quantities of interest permit such accommodation.

Kelling and colleagues (2009) stated that their data-driven approach is well suited to large-scale studies and asserted that "collecting data for hypothesis-testing analyses, at the spatial and temporal scope needed, is logistically, financially, or ethically challenging and is most likely not feasible for one individual expert" (p. 613). Large temporal and spatial scope is certainly challenging, but no more so for knowledge-driven scientific approaches (which are, incidentally, frequently conducted by teams rather than by single individuals) than for data-driven approaches (e.g., Robbins et al. 1989, Boulinier et al. 1998, 2001, Doherty et al. 2003, Karanth KU et al. 2004, Viljugrein et al. 2005, Karanth KK et al. 2010). A related claim is that omnibus monitoring programs that are not specifically directed at either prior scientific hypotheses or conservation decision problems are somehow more able to be large in geographic scale and more likely to be useful for exploratory analyses than their more-focused counterparts. Once again, this is a fallacy, since there are no limits imposed on the scale of relevant hypotheses or conservation problems or on the kinds of analyses that can be conducted on data from focused monitoring in addition to the uses for which they are designed (Nichols and Williams 2006; also see below).

Concluding comments and recommendations

We very much agree with Kelling and colleagues (2009) that there is much left to learn about ecological systems and biodiversity, and we recognize that their data-driven approach may be capable of generating useful hypotheses. Our primary sources of disagreement are about whether hypothesis generation should be an area of emphasis at this time and, if so, whether pattern-recognition approaches are really likely to be as efficient and useful in this endeavor as knowledge-based approaches that build on over a century of ecological research. These sources of disagreement would not merit this discussion if there were no real consequence to deciding to take one approach or the other to the conduct of biodiversity science. But there are consequences to this decision. From a societal perspective, funding and effort that can be devoted to conservation are severely limited, which emphasizes the importance of how we allocate our resources as conservation scientists and managers. The steps described by Kelling and colleagues (2009) to acquire, properly organize, store, and then use observational data require substantial time, effort, and expense. Therefore, questions about the relative efficiency of data-driven versus knowledge-driven approaches are important. From the perspective of the individual scientist, professional life spans—and, therefore, lifetime lists of completed projects—are finite and relatively short. Research and conservation projects typically require several years, at least, to complete. Therefore, periodic decisions about what questions to address and how to address them are among the most important that an investigator ever makes. We believe that the ideas presented here deserve consideration in such decisions.

A substantial portion of the expense and effort associated with the data-intensive approach of Kelling and colleagues (2009) is associated with the acquisition, organization, synthesis, and storage of large amounts of observational data. One recommendation for reducing the cost and effort associated with their approach would be to focus their exploratory analyses on data collected for the conduct of knowledge-driven science (Nichols and Williams 2006). We have advocated knowledge-driven research directed at discriminating among the numerous competing hypotheses that currently exist about the structure, function, and dynamics of ecological systems. This research requires analyses and, when possible, data-collection efforts that are guided by prior hypotheses. However, following inferences from such analyses, we can see a role for the strategic use of exploratory analyses directed at hypothesis generation. We believe that such exploratory analyses are most likely to be useful when they focus on hypotheses suggested by the original analysis. In particular, it seems sensible to explore new hypotheses that might explain deviations of data from the prior hypotheses being investigated, a common step in Kuhn's (1962) "normal science." We suggest that the presentation of such analyses belongs in discussion sections of scientific articles and should be clearly labeled as exploratory and directed at hypothesis generation. This recommendation would permit the data-driven analyses advocated by Kelling and colleagues (2009), but the analyses would be based on data obtained as part of a program of knowledge-based science. As such, the utility of the data for addressing prior hypotheses would be guaranteed, and the possibility of finding patterns to generate useful new hypotheses would still exist. As is emphasized in our arguments here, the complementary approach of using data collected as a component of omnibus, data-driven science for the purpose of discriminating among prior competing hypotheses is much less likely to be successful.

We acknowledge that we have not presented a detailed cost–benefit analysis for the two contrasted approaches to ecosystem science. We have strong beliefs and hopefully logical arguments about the relative benefits associated with the two discussed approaches to science, but we do not know how to adequately quantify these arguments (e.g., what is the ratio of probabilities of developing a useful new hypothesis through approaches using pattern recognition versus prior knowledge?). Therefore, we have not considered the optimal allocation of effort and resources to data-driven versus knowledge-driven approaches. Instead, we have followed the advocacy approach of Kelling and colleagues (2009), arguing that the majority of our effort and resources should be expended on knowledge-driven approaches to science but admitting the possibility of adding exploratory analyses to this approach. We ultimately agree with Fretwell's (1972) "plea for tolerance" and with his claim that "whatever works should be acceptable" (p. viii). In this article, we have simply provided our arguments about why we believe that knowledge-driven approaches should work best for us.

Acknowledgments

We benefited from constructive comments on an earlier draft of the manuscript made by Jim Hines, Jay Rotella, Mike Runge, and four anonymous reviewers.

References cited

Adam AM. 2000. Farewell to certitude: Einstein's novelty on induction and deduction, fallibilism. Journal for General Philosophy of Science 31: 19–37.

Altizer S, Hochachka WM, Dhondt AA. 2004. Seasonal dynamics of mycoplasmal conjunctivitis in eastern North American house finches. Journal of Animal Ecology 73: 309–322.

Anderson DR, Burnham KP, Gould WR, Cherry S. 2001. Concerns about finding effects that are actually spurious. Wildlife Society Bulletin 29: 311–316.

Boulinier T, Nichols JD, Hines JE, Sauer JR, Flather CH, Pollock KH. 1998. Higher temporal variability of forest breeding bird communities in fragmented landscapes. Proceedings of the National Academy of Sciences 95: 7497–7501.

———. 2001. Forest fragmentation and bird community dynamics: Inference at regional scales. Ecology 82: 1159–1169.

Brown JH. 1995. Macroecology. University of Chicago Press.

Brown JH, Stevens GC, Kaufman DM. 1996. The geographic range: Size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics 27: 597–623.

Caruana R, Elhawary M, Munson A, Riedewald M, Sorokina D, Fink D, Hochachka WM, Kelling S. 2006. Mining citizen science data to predict prevalence of wild bird species. Pages 909–915 in KDD '06: The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Philadelphia, PA, USA—August 20–23, 2006. Association for Computing Machinery.

Cole LC. 1957. Biological clock in the unicorn. Science 125: 874–876.

Conn PB, Cooch EG. 2009. Multistate capture-recapture analysis under imperfect state observation: An application to disease models. Journal of Applied Ecology 46: 486–492.

Doherty PF Jr, Sorci G, Royle JA, Hines JE, Nichols JD, Boulinier T. 2003. Sexual selection affects local extinction and turnover in bird communities. Proceedings of the National Academy of Sciences 100: 5858–5862.

Fara P. 2009. Science: A Four Thousand Year History. Oxford University Press.

Faustino CR, Jennelle CS, Connolly V, Davis AK, Swarthout EC, Dhondt AA, Cooch EG. 2004. *Mycoplasma gallisepticum* infection dynamics in a house finch population: Seasonal variation in survival, encounter and transmission rate. Journal of Animal Ecology 73: 651–669.

Fisher RA. 1947. The Design of Experiments, 4th ed. Oliver and Boyd.

Fretwell SD. 1972. Populations in a Seasonal Environment. Princeton University Press.

Gaston KJ. 2003. The Structure and Dynamics of Geographic Ranges. Oxford University Press.

Gaston KJ, Blackburn TM. 2000. Pattern and Process in Macroecology. Wiley.

Green RH. 1979. Sampling Design and Statistical Methods for Environmental Biologists. Wiley.

Hand DJ. 2005. What you get is what you want? Some dangers of black box data mining. M2005 Conference Proceedings. SAS Institute.

Hilborn R, Mangel M. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press.

Hochachka WM, Caruana R, Fink D, Munson A, Riedewald M, Sorokina D, Kelling S. 2007. Data mining for discovery of pattern and process in ecological systems. Journal of Wildlife Management 71: 2427–2437.

Hosseini PR, Dhondt AA, Dobson A. 2004. Seasonality and wildlife disease: How season birth, aggregation and variation in immunity affect the dynamics of *Mycoplasma gallisepticum* in house finches. Proceedings of the Royal Society B 271: 2569–2577.

Jennelle CS, Cooch EG, Conroy MJ, Senar JC. 2007. State-specific detection probabilities and disease prevalence. Ecological Applications 17: 154–167.

Karanth KK, Nichols JD, Karanth KU, Hines JE, Christensen NL Jr. 2010. The shrinking ark: Patterns of large mammal extinctions in India. Proceedings of the Royal Society 277: 1971–1979.

Karanth KU, Nichols JD, Kumar NS, Link WA, Hines JE. 2004. Tigers and their prey: Predicting carnivore densities from prey abundance. Proceedings of the National Academy of Sciences 101: 4854–4858.

Kelling S, Hochachka WM, Fink D, Riedewald M, Caruana R, Ballard G, Hooker G. 2009. Data-intensive science: A new paradigm for biodiversity studies. BioScience 59: 613–620.

Kuhn TS. 1962. The Structure of Scientific Revolutions. University of Chicago Press.

Lancia RA, Nichols JD, Pollock KH. 1994. Estimating the number of animals in wildlife populations. Pages 215–253 in Bookhout TA, ed. Research and Management Techniques for Wildlife and Habitats. The Wildlife Society.

Lafleur LJ. 1960. Translation of R. Descartes (1637), "Discourse on Method" and "Meditations." Liberal Arts Press.

Nichols JD. 2001. Using models in the conduct of science and management of natural resources. Pages 11–34 in Shenk TM, Franklin AB, eds. Modeling in Natural Resource Management: Development, Interpretation and Application. Island Press.

Nichols JD, Williams BK. 2006. Monitoring for conservation. Trends in Ecology and Evolution 21: 668–673.

Pirsig RM. 1974. Zen and the Art of Motorcycle Maintenance: An Inquiry into Values. Bantam Books.

Popper K. 1972. Objective Knowledge. Clarendon.

Raiffa H, Schlaifer RO. 1961. Applied Statistical Decision Theory. Graduate School of Business Administration, Harvard University.

Robbins CS, Sauer JR, Greenberg RS, Droege S. 1989. Population declines in North American birds that migrate to the neotropics. Proceedings of the National Academy of Sciences 86: 7658–7662.

Romesburg HC. 1981. Wildlife science: Gaining reliable knowledge. Journal of Wildlife Management 45: 293–313.

Runge MC, Converse SJ, Lyons JE. 2011. Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. Biological Conservation 144: 1214–1223.

Seger R. 1980. Against the wind. Against the Wind. Columbia Records.

Skalski JR, Robson DS. 1992. Techniques for Wildlife Investigations: Design and Analysis of Capture Data. Academic Press.

Stone L. 1992. Coloured noise or low-dimensional chaos? Proceedings of the Royal Society B 250: 77–81.

Sugihara G, May RM. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344: 734–741.

Thompson SK. 2002. Sampling, 2nd ed. Wiley.

Tukey JW. 1962. The future of data analysis. The Annals of Mathematical Statistics 33: 1–67.

———. 1977. Exploratory Data Analysis. Addison–Wesley.

Viljugrein H, Stenseth NC, Smith GW, Steinbakk GH. 2005. Density dependence in North American ducks. Ecology 86: 245–254.

Williams BK, Nichols JD, Conroy MJ. 2002. Analysis and Management of Animal Populations. Academic Press.

Yoccoz NG, Nichols JD, Boulinier T. 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution 16: 446–453.

James D. Nichols (jnichols@usgs.gov) and John R. Sauer are US Geological Survey biologists at Patuxent Wildlife Research Center, in Laurel, Maryland. Evan G. Cooch is in the Department of Natural Resources at Cornell University, in Ithaca, New York. Jonathan M. Nichols is a research scientist with the Naval Research Laboratory, in Washington, DC.