



Large-Scale Perturbations: Opportunities for Innovation

Stephen R. Carpenter

Ecology, Vol. 71, No. 6 (Dec., 1990), 2038-2043.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9658%28199012%2971%3A6%3C2038%3ALPOFI%3E2.0.CO%3B2-4>

Ecology is currently published by The Ecological Society of America.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/esa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Ecology, 71(6), 1990, pp. 2038–2043
© 1990 by the Ecological Society of America

LARGE-SCALE PERTURBATIONS: OPPORTUNITIES FOR INNOVATION¹

STEPHEN R. CARPENTER

*Center for Limnology, University of Wisconsin,
Madison, Wisconsin 53706 USA*

Abstract. Several approaches are discussed for statistical analysis of large-scale (and possibly unreplicated) ecological experiments. These include intervention analyses and comparisons of alternative models using Bayes' formula. Such techniques are unfamiliar to many ecologists and are not typically included in graduate curricula in ecology. I argue for increased training in these areas and for collaborations between statisticians and ecologists to develop innovative approaches to the analysis of large-scale perturbations.

Large-scale unplanned or natural experiments have played an important role in ecology and evolutionary biology for generations (Diamond 1986). Deliberate large-scale experiments have developed more recently, but have already made important contributions to basic ecology (Likens 1985). Large systems are the appropriate objects of experimentation when key consumers are wide-ranging, manipulations unavoidably have effects on neighboring areas, and/or response variables can be measured only at large scales (Likens 1985, Schindler 1987, Kitchell et al. 1988). When extrapolations to large scales from smaller scale studies are uncertain or controversial, large-scale experiments are often crucial (Likens 1985, Schindler 1987). Ecosystem experiments can be combined with smaller scale, mechanistic experiments to yield more insight than would be possible from either approach alone (Kitchell et al. 1988). Ecologists now have unprecedented opportunities to contribute to basic research and environmental problem-solving at large spatial and temporal scales (National Research Council 1988).

Ecosystem experiments pose pragmatic, logistic, analytic, esthetic, and even legal challenges. The most severe obstacles arise from lack of appropriate sites and adequate, sustained funding. Among ecologists, much discussion has centered on statistical issues of design, analysis, and interpretation of ecosystem experiments. A tension exists between the need to replicate and the need to study processes at appropriately large scales. This Special Feature addresses these issues and some promising approaches to statistical analysis of ecosystem experiments. I hope that these articles

will stimulate innovative thinking about ecosystem experimentation, and productive collaborations among statisticians and ecologists.

WHY REPLICATED EXPERIMENTATION IS POWERFUL

All ecologists agree that experiments should be replicated whenever possible. Genuine replicates are independent in the sense that the outcome of a given replicate has no effect on the outcome of any others. They represent the total variability affecting replicates at specified experimental conditions (Box et al. 1978: 319). Identifying genuine replicates is often troublesome for experimenters. Hurlbert (1984) provided ecologists with an excellent, thorough discussion of genuine replication. Millard (1987) and R. A. Carpenter (1989) have decried the fact that environmental monitoring data are often not replicated even when replication is possible.

Replication improves experiments in three major ways. First, replication permits accurate estimation of experimental error. Second, results become increasingly precise as the number of replicates increases. Third, replication expands the range of experimental units studied. The population from which the replicates are drawn determines the extent to which results can be generalized. In this sense, replication (in the form of repetition of key experiments by others) has been conventional practice in science for far longer than statistics has been.

The papers presented here are not intended to be an argument against replication. Rather, we pose alternatives for large-scale experiments where standard replicated designs are not practical or are impossible.

¹ For reprints of this Special Feature, see footnote 1, page 2037.

WHAT CAN WE DO WHEN REPLICATION IS IMPOSSIBLE?

Large-scale experiments, whether planned or unplanned, are frequently not replicable. The serendipitous nature of unplanned experiments usually precludes replication, and even when manipulations are planned, experimental systems are usually limited in number. Candidate systems may be so different ecologically that they do not constitute reasonable replicates. Funding levels and logistic limitations often preclude replication.

Faced with these constraints, experimenters may be tempted to use minimal replication (e.g., duplicates). However, insufficient replication may be worse than no replication at all, if the experimenter does not consider the power of the statistical tests when interpreting results. The variability of community and ecosystem variates may be so great that experiments with only two replicates cannot detect perturbation effects unless they are very large. Modest effects, even ones that are real, are not likely to be detected in such experiments (S. R. Carpenter 1989). The resources that could be invested in duplicates might be better spent on more detailed mechanistic analyses conducted within the context of the large-scale experiment (Frost et al. 1988).

One logical and valid response to these difficulties is to abandon statistics altogether. By using substantial and sustained manipulations, ecosystem experimenters have caused changes that are so large that they are convincing in the absence of statistical analysis (Hurlbert 1984). The intrinsic variability of ecosystems may be so large that rather powerful manipulations would be needed to detect responses even if experiments could be replicated (S. R. Carpenter 1989).

I will argue, on the other hand, that there are compelling reasons to pursue statistical analysis of large-scale perturbations. The less pronounced responses of communities and ecosystems to perturbation may be the most valuable indicators of stress or impending change. Schindler (1987) has made this point forcefully:

Most of the well-studied perturbations that have severely disrupted ecosystem or community organization or function have been "sledgehammer blows." . . . Such studies are useful, but it would be risky to rely on them alone to predict responses of natural ecosystems to slowly increasing low-level perturbations. The dilemma we face is simply expressed: how do we identify when critical, measurable variables begin to vary outside the normal range, thereby indicating that the ecosystem is perturbed or stressed? (Schindler 1987:11)

Statistical analysis could be enormously valuable for the identification of change in such indicator variables.

At its best, statistical analysis sharpens thinking about data, reveals new patterns, prompts creative thinking, and stimulates productive discussions in multidisciplinary research groups. For many scientists, these positive possibilities of statistics are overshadowed by negatives: abstruse assumptions, emphasis of things one can't do, and convoluted logic based on hypothesis rejection. One colleague's reaction to this Special Feature was that "statistics is the scientific equivalent of a trip to the dentist." This view is probably widespread. It leads to insufficient awareness of the fact that statistics, like ecology, is a vital, evolving discipline with ever-changing capabilities. Most ecological experimentation is rooted in statistical methods developed for agriculture >50 yr ago. However, the statistical approaches that show promise for large-scale experiments derive from more recent developments in statistics that are not well known to ecologists (Frost et al. 1988:250).

The question of inference in large-scale experiments has two fundamental components: Are the differences between the premanipulation and postmanipulation periods nonrandom, and did the manipulation cause the differences (Frost et al. 1988:248–253). The next two sections of this paper will briefly summarize methods for detection of nonrandom change and attribution of cause in large-scale experiments.

DETECTION OF NONRANDOM CHANGE

One of the earliest statistical analyses of a large-scale perturbation clearly illustrates the detection of nonrandom change in an unreplicated experiment. Box and Tiao (1975) considered the possible effects of two events on the concentration of atmospheric ozone in downtown Los Angeles (Fig. 1). Event A was diversion of traffic away from the city by a new freeway, combined with a new law which reduced the amount of reactive hydrocarbon in gasoline. Event B was a law requiring engine design changes in new cars. Box and Tiao asked if the apparent changes following events A and B could be explained by random processes. They present time series models for step changes (as might be expected following event A) and gradual changes (as might be expected following event B). They also discuss models for other types of interventions and responses, including "step" and "pulse" interventions that correspond to the "press" and "pulse" experiments of Bender et al. (1984).

For the Los Angeles ozone data, a time series model was fit with terms for a step change occurring at event A, summer and winter gradual changes following event B, a moving average (which represents the dependency of sequential observations), and a seasonal effect. Summer and winter changes were modeled separately because the effects of engine changes after event B were expected to depend seasonally on solar radiation and

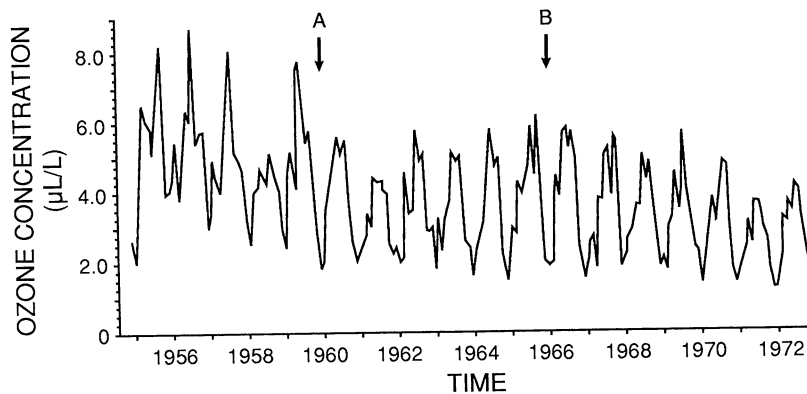


FIG. 1. Concentration of ozone in downtown Los Angeles vs. time, 1956–1972. Data shown are monthly means of hourly observations. Events A and B were interventions, explained in *Detection of nonrandom change*, which were expected to reduce ozone concentrations. Adapted from Box and Tiao (1975) with permission.

atmospheric temperature inversions. Noting that residuals were nearly normally distributed and showed no obvious deficiencies in the model, Box and Tiao drew their conclusions from the parameter estimates (Table 1). Following event A, there was a step reduction in ozone concentration of $1.09 \mu\text{L}/\text{L}$. Following event B, there was a gradual reduction of $0.25 \mu\text{L}\cdot\text{L}^{-1}\cdot\text{yr}^{-1}$ during summer, but no appreciable decrement during winter. Most important, they concluded that these reductions were nonrandom.

The analysis by Box and Tiao does not prove that the intervening events caused the nonrandom changes (Box et al. 1978:604). This point is critical, because conventional statistical tests in adequately replicated, properly randomized experiments can show whether manipulation caused changes in response variables. In Box and Tiao's intervention analysis, whether the reductions in ozone concentration were caused by the interventions must be decided on nonstatistical grounds. The intervention analysis makes a useful contribution in showing that noise in the time series is not likely to explain the ozone reductions. It also provides quantitative measures of the effects of the interventions.

Numerous approaches exist for determining whether changes following manipulations are nonrandom. The general approach proposed by Box and Tiao (1975) is flexible enough to apply to many large-scale ecological experiments, although long time series are required to fit the models. Some alternative approaches are described briefly below. Mohr (1988) provides a thoughtful and thorough discussion of analysis and inference in experiments that do not fit the strictures of conventional statistics.

Many large-scale ecological experiments have the advantage of an untreated reference (or "control") area (Likens 1985). The reference and manipulated systems are sampled in parallel by the same methods before

and after manipulation. Changes that occur in both reference and manipulated systems could be due to some factor other than the manipulation. On the other hand, nonrandom changes that occur in the manipulated system but not the reference system can be attributed to the manipulation. Premanipulation sampling is essential to determining whether the intersystem difference changed after manipulation. Statistical approaches to experiments with manipulated and reference ecosystems compare the intersystem difference before manipulation to the intersystem difference after manipulation (Stewart-Oaten et al. 1986, Carpenter et al. 1989). Published approaches deal with step changes only. Considerably greater flexibility could be gained by applying the more general approach of Box and Tiao (1975) to time series of intersystem differences.

The advent of cheap, rapid computation makes feasible the determination of error distributions by random permutation of the data (Diaconis and Efron 1983). These procedures eliminate the need to assume that data follow particular distributions, such as the normal distribution. Carpenter et al. (1989) proposed and evaluated a randomization test for ecosystem experiments. Conclusions based on the randomization test were essentially the same as those based on conventional t

TABLE 1. Maximum likelihood estimates (MLE) and standard errors (SE) of parameters fit by Box and Tiao (1975) to the time series of Fig. 1.

Parameter	MLE	SE
Step change after event A	-1.09	0.13
Rate of annual decline in summer ozone concentration after event B	-0.25	0.07
Rate of annual decline in winter ozone concentration after event B	-0.07	0.06
Moving average term	-0.24	0.03
Seasonal term	0.55	0.04

tests. Edgington (1980) has provided a practical guide to randomization tests, but readers should be aware of theoretical errors pointed out by Green (1981) and Good (1980). Randomization tests for multivariate ecological data, such as community censuses, have been developed and applied to large-scale community changes (LeGendre and Fortin 1989).

Sometimes it is possible to compare a single manipulated ecosystem with many reference systems using conventional statistics (Schindler et al. 1985, Schindler 1987, Carpenter et al. 1989). For example, a conventional *t* test was used to compare a manipulated lake with nine reference lakes (Carpenter et al. 1989). The advantage of this approach is stronger inference: a small *P* value implies that the manipulation caused the deviation of the manipulated system. However, the capacity of such tests to detect real changes is much less than that of a test with the same number of ecosystems allocated equally to manipulated and reference groups (Carpenter et al. 1989).

CAUSAL INFERENCE

If nonrandom changes occur, one must then ask whether the manipulation or some other factor is the most likely cause of the changes (Frost et al. 1988). Often, ecological criteria, not statistical ones, are the basis for evaluation (Frost et al. 1988, Carpenter et al. 1989). A quantitative analysis is possible when alternative hypotheses can be compared using Bayesian statistics. Bayesian analyses are widely used in natural resource management and environmental engineering (Reilly 1970, Blau and Neely 1975, Walters 1986, Reckhow 1990, Walters and Holling 1990), but are not yet common in community and ecosystem ecology (Carpenter et al. 1990).

Bayesian statistics can be used to compare several alternative hypotheses, or models, in a common framework. The analysis involves a data set (*Y*) and a set of candidate models (M_i). The models are chosen to represent distinct alternative explanations, mechanisms, or policy options. Walters (1986) discusses in detail the selection of model sets. Prior to analysis, one assigns a prior probability $P_i(0)$ that each model is correct. The prior probabilities may be calculated from previous studies. If no prior information is available, prior probabilities may be assigned subjectively. Subjective assignment of prior probabilities is often criticized. However, in most applications the prior probabilities have far less influence than the data on the outcome of the analysis. In any case, one can easily calculate the effects of the prior probabilities on the conclusions. For a detailed discussion of prior probabilities and their effects, see Box and Tiao (1973). Once the data, model set, and prior probabilities are in hand, one computes the likelihood of obtaining the data given

each model, $L(Y|M_i)$. The likelihood usually is derived from the residuals after least-squares fitting of the model. For normally distributed residuals, the probability *p* of any residual *w* is

$$p = \exp(-w^2/2s^2)/(2\pi s^2)^{0.5},$$

where *s* is the standard deviation of all the residuals. The likelihood *L* is simply the product of the probabilities for all the residuals. Since the residuals are centered around zero, this product simplifies to

$$L = \exp(-N/2)/[(2\pi s^2)^{0.5}]^N.$$

L decreases steeply and nonlinearly as the standard deviation of the residuals increases. The posterior probability $P_i(1)$ for each model is

$$P_i(1) = L(Y|M_i) \cdot P_i(0)/T,$$

where *T* is simply the sum of $L \cdot P_i(0)$ for all the models considered. The $P_i(0)$ are the probabilities that each model is correct, given the information in the observations *Y*. Future experiments and/or management actions derive from the posterior probabilities. In scientific applications, the usual goal is to find one best model. Techniques exist for designing experiments that maximize the probability of discriminating among the models (Walters 1986).

To illustrate the Bayesian approach, I have expanded an example by Blau and Neely (1975). They sought a model for flux of a pesticide through a pond ecosystem. In a whole-system experiment using radioactively labeled pesticide, most of the label was associated with water, sediments, and fishes. For each of eight plausible explanations of ecosystem response, a model was erected (Table 2). Blau and Neely represented each model (Table 2) as a system of differential equations. For the Bayesian analysis, I assumed that each model had an equal prior probability of 1/8. Posterior probabilities were calculated using likelihood values provided by Blau and Neely. Five of the seven models have extremely small posterior probabilities and are effectively eliminated from further consideration. Model 4a is by far the most likely one.

CONCLUSIONS

When ecological questions require large-scale experimentation, unreplicated experiments can be extremely informative. Several statistical analyses, unfamiliar to many ecologists, are available to guide the interpretation of unreplicated experiments. Some of these techniques are discussed in this Special Feature; for other possibilities, see Frost et al. (1988). While genuine replication is a powerful tool that should be used when possible, the scale of ecological research should not be dictated by statistical constraints.

TABLE 2. Features of models of chemical transfer among water, sediment, and fish considered by Blau and Neely (1975), with likelihoods calculated by them and posterior probabilities calculated by me under the assumption of equal prior probability. Model numbers correspond to those in their paper.

Feature	1	2a	2b	3a	3b	4a	4b
Partitioning of chemical between sediment and water	X	X	X	X	X	X	X
Direct uptake of chemical by fish	X	X	X	X	X	X	X
Chemical excreted unchanged by fish	X						
Chemical modified, then excreted by fish		X	X	X	X	X	X
Modified, excreted chemical taken up directly by sediment				X		X	X
Modified, excreted chemical partitions between water and sediment					X		
Chemical in fish partitions between two tissues (e.g., lipid, nonlipid)						X	
Modified, excreted chemical partitions between fish and water							X
Log likelihood	-384	-187	-56.4	-10.1	-10.0	0.758	-0.745
Posterior probability	←—————>			<10 ⁻⁸		—————>	
						0.97	0.03

We will learn more by repeating certain key ecosystem experiments than we would have by replicating the original experiment. Testing ideas in a broader range of systems, one of the most important benefits of replication, can also be achieved by repetition. Repetition, in a different region, by a different research team that may use different methods, broadens our experience far more than mere replication. Empirical Bayesian analysis even allows quantitative conclusions from combined results of different studies (Morris 1983). Concern with time-treatment interaction (Walters and Holling 1990) underscores the importance of repeating key experiments.

The opportunities for new approaches discussed in this Special Feature contrast sharply with the statistical training provided to most graduate students in ecology. Ecology graduate students typically learn statistics designed for laboratory containers and small test plots, and/or multivariate techniques used for describing communities and ecosystems. Experimental science at the scale of communities and ecosystems needs graduate students trained in time series analysis, Bayesian approaches, and computer-intensive statistics.

The statistical challenges of large-scale experimentation appear less daunting than the practical ones. Effective ecosystem experiments require substantial funding sustained for years or even generations of researchers. Sites for large-scale experiments are rare. The value of dedicated tracts with long-term monitoring of multiple reference-systems is amply demonstrated by Canada's Experimental Lakes Area (Schindler 1987). Long-term data for estimating baseline variability are also uncommon, though the Long-Term Ecological Research program will help fill that gap. We need more opportunities for important ecosystem experiments to be performed, replicated or repeated, and coordinated with long-term records.

ACKNOWLEDGMENTS

I thank Jim Kitchell and Tom Frost for extensive discussions of these issues. They, Pam Matson, and Gary Sprules provided helpful reviews of the manuscript. Jill Coburn prepared the manuscript and Cheryle Hughes drafted the figure. I'm grateful for the support of NSF grant BSR-86-04996.

LITERATURE CITED

- Bender, E. A., T. J. Case, and M. E. Gilpin. 1984. Perturbation experiments in community ecology: theory and practice. *Ecology* **65**:1-13.
- Blau, G. E., and W. B. Neely. 1975. Mathematical model building with an application to determine the distribution of Dursban insecticide added to a simulated ecosystem. In A. MacFayden, editor. *Advances in ecological research*. Volume 9. Academic Press, London, England.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for experimenters*. John Wiley and Sons, New York, New York, USA.
- Box, G. E. P., and G. C. Tiao. 1973. *Bayesian inference in statistical analysis*. Addison-Wesley, Reading, Massachusetts, USA.
- Box, G. E. P., and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental parameters. *Journal of the American Statistical Association* **70**:70-79.
- Carpenter, R. A. 1989. Do we know what we are talking about? *Land Degradation and Rehabilitation* **1**:1-3.
- Carpenter, S. R. 1989. Replication and treatment strength in whole-lake experiments. *Ecology* **70**:453-463.
- Carpenter, S. R., T. M. Frost, D. Heisey, and T. Kratz. 1989. Randomized intervention analysis and the interpretation of whole ecosystem experiments. *Ecology* **70**:1142-1152.
- Carpenter, S. R., T. M. Frost, J. F. Kitchell, T. K. Kratz, D. W. Schindler, J. Shearer, W. G. Sprules, M. J. Vanni, and A. P. Zimmerman. 1990. Patterns of primary production and herbivory in 25 North American lake ecosystems. In J. Cole, S. Findlay, and G. Lovett, editors. *Comparative analyses of ecosystems: patterns, mechanisms, and theories*. Springer-Verlag, New York, New York, USA, *in press*.
- Diaconis, P., and B. Efron. 1983. Computer-intensive methods in statistics. *Scientific American* **248**:116-130.
- Diamond, J. 1986. Overview: laboratory experiments, field experiments, and natural experiments. Pages 3-22 *in* J.

- Diamond and T. J. Case, editors. *Community ecology*. Harper and Row, New York, New York, USA.
- Edgington, E. S. 1980. *Randomization tests*. Marcel Dekker, New York, New York, USA.
- Frost, T. M., D. L. DeAngelis, S. M. Bartell, D. J. Hall, and S. H. Hurlbert. 1988. Scale in the design and interpretation of aquatic community research. Pages 229–260 *in* S. R. Carpenter, editor. *Complex interactions in lake communities*. Springer-Verlag, New York, New York, USA.
- Good, P. 1980. Book review: randomization tests. *Journal of Statistical Computing and Simulation* **11**:157–160.
- Green, B. F. 1981. Randomization tests. *Journal of the American Statistical Association* **76**:495.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187–211.
- Kitchell, J. F., S. M. Bartell, S. R. Carpenter, D. J. Hall, D. J. McQueen, W. E. Neill, D. Scavia, and E. E. Werner. 1988. Epistemology, experiments, and pragmatism. Pages 263–280 *in* S. R. Carpenter, editor. *Complex interactions in lake communities*. Springer-Verlag, New York, New York, USA.
- LeGendre, P., and M. J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio*, *in press*.
- Likens, G. E. 1985. An experimental approach for the study of ecosystems. *Journal of Ecology* **73**:381–396.
- Millard, S. P. 1987. Environmental monitoring, statistics, and the law: room for improvement. *American Statistician* **41**:249–253.
- Mohr, L. B. 1988. *Impact analysis for program evaluation*. Dorsey, Chicago, Illinois, USA.
- Morris, C. 1983. Parametric empirical Bayes' inference: theory and applications. *Journal of the American Statistical Association* **78**:47–65.
- National Research Council. 1988. *Toward an understanding of global change*. National Academy Press, Washington, D.C., USA.
- Reckhow, K. H. 1990. Bayesian inference in non-replicated ecological studies. *Ecology* **71**:2053–2059.
- Reilly, P. M. 1970. Statistical methods in model discrimination. *Canadian Journal of Chemical Engineering* **48**:168–173.
- Schindler, D. W. 1987. Detecting ecosystem response to anthropogenic stress. *Canadian Journal of Fisheries and Aquatic Sciences* **44**(Supplement):6–25.
- Schindler, D. W., K. H. Mills, D. F. Malley, D. L. Findlay, J. A. Shearer, I. J. Davies, M. A. Turner, G. A. Linsley, and D. R. Cruikshank. 1985. Long-term ecosystem stress: the effects of years of experimental acidification on a small lake. *Science* **228**:1395–1401.
- Stewart-Oaten, A., W. Murdoch, and K. Parker. 1986. Environmental impact assessment: "pseudoreplication" in time? *Ecology* **67**:929–940.
- Walters, C. J. 1986. *Adaptive management of renewable resources*. Macmillan, New York, New York, USA.
- Walters, C. J., and C. S. Holling. 1990. Large-scale management experiments and learning by doing. *Ecology* **71**:2060–2068.