CONSERVATION ISSUES

The Use of Control Charts to Interpret Environmental Monitoring Data

Lloyd W. Morrison[1]

Heartland Inventory and Monitoring Network
National Park Service
and
Department of Biology
Missouri State University
901 S. National Avenue
Springfield, MO 65897

[1] Corresponding author: LloydMorrison@MissouriState.edu; 417-836-3119

Natural Areas Journal 28:66–73

ABSTRACT: Many different methods of synthesizing and analyzing environmental monitoring data exist. Given the diversity of current environmental monitoring projects, and the large number of scientists and policy-makers involved, there is a critical need for a universal format that both summarizes data sets and indicates any potential need for management action. Control charts, originally developed for industrial applications, represent one way of doing this. Control charts indicate when a system is going 'out of control' by plotting through time some measure of a stochastic process with reference to its expected value. Control charts can be constructed for many different types of indicators, whether univariate or multivariate. Control charts are simple to interpret, and can easily be updated whenever additional data become available. The relative risks of Type I (i.e., concluding meaningful change has occurred when actually it has not) and Type II (i.e., concluding meaningful change has not occurred when in fact it has) errors are intuitive and easily adjusted, and one may define a threshold for action at any desired level. Control charts may often be more informative than traditional statistical analyses such as regressions or parameter estimation with confidence intervals. The primary challenge in most situations will be determining a stable or baseline state for the ecological indicator in question.

Index terms: confidence intervals, control chart, control limit, ecological indicator, environmental monitoring

INTRODUCTION

Monitoring the ecological integrity of natural resources is an important activity for many government agencies, university researchers, and private foundations concerned with conservation. As the world changes due to human-induced pressures and natural sources of variation, there is an increasing need to document these changes and take action, when necessary and possible, to halt or reverse undesirable trends. Frequently many individuals will be involved in this process for any variable of interest. For example, some will be involved in data collection, others in data analysis, still others in interpretation, and a final group charged with making decisions based on the results.

The type of information needed may also vary; for example, at times one may be interested in a measure of central tendency (e.g., a mean) and at other times a measure of variability (e.g., a range). Given the large number of response variables that are currently being evaluated (or are planned for evaluation) and the plethora of statistical (and astatistical) methods available for such analyses, the ultimate number of different types of data organization or analysis 'products' is very large. For example, data from monitoring projects could be analyzed by different types of regression approaches, ANOVAs, time series analyses, Monte Carlo modeling, plotted as simple graphs without the use of inferential statistics, etc. (e.g., Hatfield et al. 1996; Thomas 1996; Dixon et al. 1998; Elzinga et al. 2001).

Ultimately, the critical component of this entire process rests with those charged with making decisions based on interpretation of data. These are frequently resource managers who may not have an in-depth knowledge of statistics or busy administrators who may devote only a few minutes to any given issue. The production of many varied types of data summaries and analyses, some of which may be quite complicated, may have the effect of creating confusion and indecision. Thus, there exists a critical need for some type of 'common currency' for the process of data analysis and reporting in environmental monitoring programs. We need a universal format that summarizes the data and indicates any potential need for management action. Ideally, this format would be the same for any type of variable. It would be straightforward to interpret, simple to update, and easy for someone with little time and without a strong statistical background to obtain a basic, yet accurate understanding of the issue.

The solution lies in the use of control charts, originally developed for, and used frequently in, industrial applications. Control charts indicate when a system is going 'out of control' by plotting through time some measure of a stochastic process with reference to its expected value. Control charts may be univariate or multivariate, representing one or more than one 'quality characteristic.'

CONTROL CHART FUNDAMENTALS

Control charts have a long history in industry as a component of statistical process control (e.g., Wetherill and Brown 1991; Beauregard et al. 1992; Gyrna 2001; Montgomery 2001). The theoretical basis for a control chart is illustrated in Figure 1. The value of some 'quality characteristic' is plotted along the *y*-axis, whereas the *x*-axis represents time or sample number. There is a target or centerline value for an 'in-control' process. The actual value is expected to vary randomly around this centerline over time. Upper and lower 'control limits' specify thresholds beyond which variability in the quality characteristic indicates a process is 'out of control.' This type of control chart is known as a Shewhart control chart (Shewhart 1931).

Control charts may be constructed for numerous variables of interest, including measures of central tendency and variability. In manufacturing, control charts are also constructed for attribute or count data. In this case, the quality characteristic would represent a proportion or number.

Control charts have been suggested for use in natural resource monitoring (e.g., (McBean and Rovers 1998; Manly 2001), although such application of control charts appears very limited (see Atkinson et al. 2003 for an example). In this paper, I focus on the application of control charts to natural resource monitoring, and employ somewhat different terminology than traditionally used in industry. I use the term 'indicator' in place of quality characteristic, to represent the natural resource of interest. I also employ the term 'control limit' more broadly, used here to represent any threshold at which management should be alarmed and consider action, regardless of the statistical basis of this threshold.

Most traditional control charts assume that observations come from a normal distribution or that data can be transformed to normality. In industry, control limits are often set at a distance of three standard deviations on either side of the centerline (Wetherill and Brown 1991; Beauregard et al. 1992; Montgomery 2001). Thus, assuming a normal distribution centered at the centerline, the control limits would encompass 99.73% of the distribution.

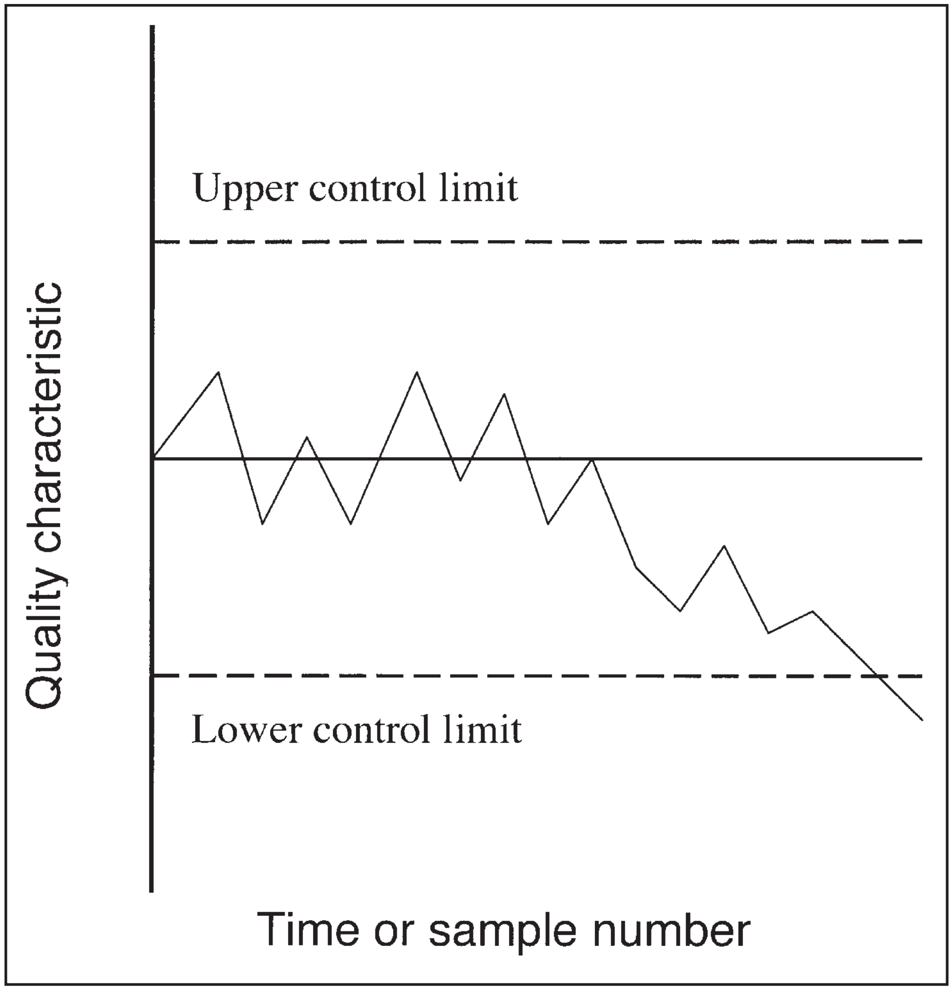

Figure 1. Theoretical basis for a control chart. In industry, the centerline value represents an 'in-control' process, which is analogous to a baseline or 'normal' (or target) value for ecological indicators.

Control limits may be constructed to contain any desired proportion of the distribution (i.e., representing [1-α] confidence intervals for any α). In this case, choosing control limits is equivalent to specifying a critical region for testing the hypothesis that a specific observation is statistically different from the proposed centerline value. (It is crucial that the centerline value is representative of the true population parameter.) Control limits could also be based on probabilistic thresholds other than confidence intervals (e.g., McBean and Rovers 1998).

If the observations cannot be assumed to come from a normal distribution, there are several options available beyond simple transformations of data. One option is to create subgroups of consecutive samples and then use the subgroup averages, which will be approximately normally distributed in accordance with the central limit theorem (see Beauregard et al. 1992; Montgomery 2001). It is possible to construct control charts based on other distributions (e.g., a Poisson distribution as in Atkinson et al. 2003) and construct analogous confidence limits, as long as the distributions are known. Distribution-free confidence limits may also be calculated, although these will usually be relatively wide and less sensitive to changes (Conover 1999).

It is not absolutely necessary to use values from a statistical sampling process to determine centerlines and thresholds for action. It is possible to subjectively choose a centerline value as the desired state and set threshold limits to match the amount of variability with which one is comfortable for the variable of interest. It is crucial to realize that this approach has no statistical basis, and thus probabilities cannot be readily associated with the observations. This application also has a precedent in industry. Such charts, which plot observations without relevance to an underlying distribution, have been termed 'conformance charts.' Threshold values, which may be subjective, are termed 'action limits' (Beauregard et al. 1992). If taking this approach in the context of environmental monitoring, one should be very familiar with the system in question and, preferably, select values that are defensible based on scientific data.

MULTIVARIATE CONTROL CHARTS

Most applications of control charts in industry have been of the univariate type. In environmental monitoring, however, we often simultaneously track multiple, inter-related variables, and frequently large, complex communities. Moreover, methods employing multivariate measures of species assemblages may be more sensitive to change than single indicator species or univariate indices (e.g., the Shannon diversity index) (Clarke 1993; Pettersson 1998). How does one produce a simple control chart that simultaneously summarizes multiple variables?

In industry, the Hotelling T^2 multivariate control chart has traditionally been used for simultaneous evaluation of two or more variables of interest. It is based on a chi-square statistic, and thus requires a multivariate normal distribution. Such conventional multivariate control charts are useful when the number of variables is relatively small (<10). When a relatively large number of variables are of simultaneous interest, the traditional approach has been to reduce the dimensionality of the problem, usually through principal components analysis (Montgomery 2001).

A new type of multivariate control chart, however, has recently been proposed for use with complex ecological communities (described in Anderson and Thompson 2004), and appears to have utility for long-term environmental monitoring. In contrast to other approaches, this method does not require any specific distributions of variables. In general, species abundances are not distributed as multivariate normal (Taylor 1961; Gaston and McArdle 1994). Traditional multivariate procedures are frequently not robust to violations of this assumption (Mardia 1971; Olson 1974).

The method requires species abundance data from one or more sites collected over a number of years. The data could represent species numbers, frequencies, biomass, etc. There is no limit to the number of species that could be included. Principle coordinate analysis (a.k.a. metric multidimensional scaling) is applied to the species abundance data. In theory, any distance or dissimilarity measure could be used, although some may have advantages over others (Anderson and Thompson 2004). This produces an ordination in which a point from each observation period is plotted in multivariate *p*-space through time. The first two resulting coordinates are used to construct the control chart. The y-axis in the control chart represents the distance between two points in the ordination graph. The deviation of a particular observation can be based on the distance from each observation to a centroid based on all previous observations or from a mean calculated from a baseline set of observations. The user specifies the baseline period. A bootstrapping procedure may be used to obtain control limits, which can be set at any desired value. Only upper control limits are relevant, because all plotted values will be distances, which are positive. The final product is similar to a univariate control chart (Figure 2).

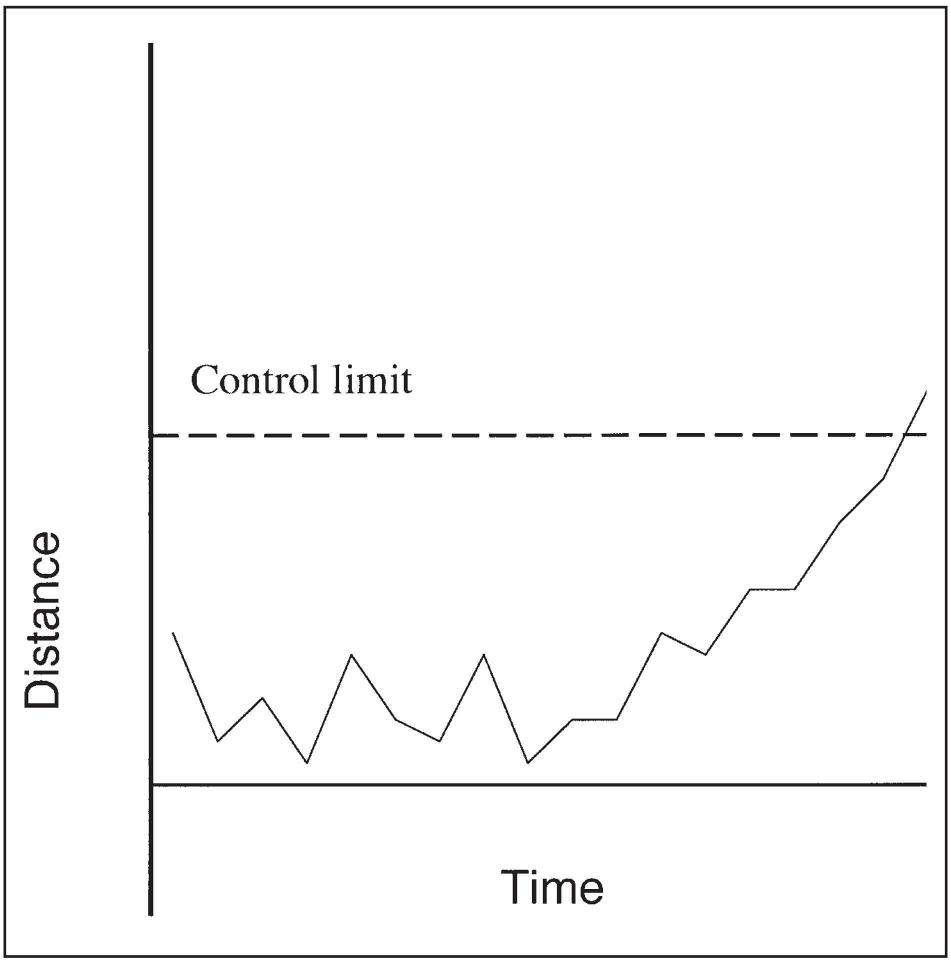

Figure 2. The basic elements of a multivariate control chart, based on the method of Anderson and Thompson (2004). The *y*-axis indicates the distance between the most recent observation and a 'baseline' average position of all or some previous observations, as determined by plotting the first two coordinates determined by principle coordinate analysis. The control limit is determined by a bootstrapping procedure. Because all distances are positive, only an upper control limit is needed.

DETERMINING THE CENTERLINE VALUE

There is a fundamental difference between the use of control charts in industrial processes compared with natural resources monitoring. In industry, an engineer usually determines the centerline for the quality characteristic, and the machinery is designed and adjusted accordingly. Relatively little variability exists around the centerline, which may be known before any samples are taken. In natural resources monitoring, we are dealing with parameters that are often unknown (before the commencement of monitoring) and are always changing, at least over some time scale. Relatively large variabilities frequently characterize environmental data. Issues of seasonality and phenology may also complicate the situation.

Thus it will usually be more difficult to determine a 'normal' or baseline state for any environmental indicator, whether univariate or multivariate, relative to industrial applications. In some cases, we may have long-term datasets that provide a basis for determining the centerline. In other cases, however, we may have relatively little or no such data. The baseline value may be determined from a relatively short time period, although this baseline may be unstable, and re-evaluation of the robustness of this estimate may be necessary over time as more data become available. Another option would be to combine empirical data with a subjective estimate of the desired state of the resource to determine a centerline value and control limits. This would be a somewhat 'Bayesian' approach, and would make calculation of probabilities difficult (or impossible). Finally, one may simply employ a set of 'desired' conditions (as in conformance charts in industry), although any thresholds would not be statistically based. Preferably, such decisions would be based on knowledge of the natural history of the system in conjunction with comparable data from other systems.

How many observations are necessary to establish the baseline period? For industrial applications, something on the order of 20 samples is usually recommended (Beauregard et al. 1992; Montgomery 2001). In environmental monitoring scenarios, it may not be practical to wait this long before establishing a baseline, especially if samples are collected annually, and baselines may need to be established based on a smaller number of samples (e.g., Atkinson et al. 2003). In any case, there are no substitutes for good judgment. In some cases, it may be informative to evaluate multiple centerlines (e.g., Anderson and Thompson 2004).

INTERPRETATION

The most basic interpretation of a control chart is that the process is in control when observations fall within the control limits and out of control when observations exceed these limits. There are many other types of patterns, however, that are indicative of a smaller degree of change in the indicator over time. In fact, any non-random pattern may be indicative of an important change. For example, if one assumes a normal distribution, 50% of the observations would be expected to fall on either side of the centerline. If a number of consecutive observations fell on one side of the centerline, this would be cause for concern. A number of other recognized patterns that indicate a departure from randomness in control charts are described by Beauregard et al. (1992) and Montgomery (2001).

In a long-term environmental monitoring program, it must be realized that a single observation falling outside a control limit does not necessarily represent an out of control process. In the case of control limits representing (1-α) confidence intervals, it would be expected that α of the observations would fall within the control limits. Thus, one out of every 1/α observations would be expected to fall outside these limits based on natural variability in the system, independent of any systematic change. (For example, if $\alpha = 0.05$, 1 out of every 20 observations would fall outside the limits.) Thus, an occasional observation above or below the control limits may not be cause for alarm, particularly if there are an equal number of observations falling above and below the upper and lower control limits, respectively.

As in more traditional statistical approaches, control charts allow for determination of the likelihood that certain observations are due to chance variability, rather than an actual trend. In traditional null hypothesis significance testing (NHST), this is known as evaluation of the Type I error rate. A Type I error in this case would represent change, concluding that one has found a significant change when there was none. Control charts allow for a very intuitive approach to evaluating the likelihood of such 'false positives.' For example, for a particular indicator we may assume a normal distribution (setting the centerline at the center of this distribution) and calculate 95% confidence intervals to serve as control limits. The likelihood of various observations falling beyond a given control limit could be easily calculated.

The other type of mistake one could make when evaluating such data would be to conclude that the population is not changing when in fact it is (i.e., a Type II error). Evaluation of the Type II error rate is known as power analysis. Power is often relatively low in many NHST approaches to analyzing environmental monitoring data, especially during the first decade or so of monitoring (e.g., Gibbs and Melvin 1997; Hayes and Steidl 1997; Van Strien et al. 1997; Gibbs et al. 1998). In environmental monitoring, it may often be desirable to control primarily for the Type II error rate, rather than the Type I error rate, which is what is usually done (i.e., α is usually specified in the statistical analysis, and power will be a function of α and other variables) (Lindley et al. 2000).

Adjustment and interpretation of both types of error are simple and straightforward in the use of control charts. Widening the control limits decreases the risk of a Type I error, and increases the risk of a Type II error. The opposite is true if the limits are narrowed.

COMPARISON WITH OTHER METHODS

The use of control charts has a number of advantages compared to more traditionally employed methods of evaluating environmental monitoring data. For example, regression analyses are frequently used in attempts to detect significant trends. The pattern of change in the parameter of interest may not be linear, however, nor very nearly approximate other commonly used regression functions. Moreover, many years may be required before enough data points are obtained so that a regression becomes 'significant,' as the *P*-value in such analyses is strongly influenced by sample size. The fewer observations, the less likely the relationship will be deemed significant, even if a real trend is occurring (Utts 1988; Johnson 1999; Anderson et al. 2000). Greater natural variability in the variable of interest (more scatter of the points around the regression function) will extend this time even farther. In contrast, it is possible to obtain a 'significant' result from a control chart in a shorter period of time.

Most statistical methods of testing for a trend employ a NHST framework. NHST has been criticized, both in general and specifically as used for ecological data. We know that no two populations will be exactly the same at different points in time. If we fail to find a significant difference, it is simply because our sample size was too small or we were not able to measure the population accurately enough (or the population was too variable). Thus, our null hypothesis of no change is trivial, and known to be false before we begin (Cherry 1998; Johnson 1999; Anderson et al. 2000). Many statisticians have advocated the use of parameter estimation along with confidence intervals instead of NHST (e.g., Carver 1993; Kirk 1996; Hoenig and Heisey 2001; Colegrave and Ruxton 2003; Nakagawa and Foster 2004).

Are control charts considered parameter estimation or null hypothesis significance testing? They contain elements of both. In a 'pure' parameter estimation approach, one would estimate the value of the parameter of interest and construct a confidence interval for each point in time, and then try to determine which points were different from each other. This is not always a straightforward task, however, and recent studies have shown that confidence intervals are frequently misinterpreted (Cumming et al. 2004; Belia et al. 2005).

With a control chart approach, one constructs a control limit (which may represent a confidence interval or some threshold of biological significance) and compares succeeding observations to this limit. As such, one may be testing the hypothesis that an observation is different from the centerline value for each new observation, depending upon how control limits are defined. Thus, whether this approach qualifies strictly as a traditional null hypothesis significance test may be debated, and would vary from one application to another.

Unlike a traditional NHST, exact *P*-values are not reported with control charts (although they could be calculated for each observation, assuming the appropriate distribution and what the control limits represent). The problem with *P*-values is that the 0.05 threshold is often interpreted as an absolute difference between night and day, regardless of whether *P* is 0.06 or 0.9 (Yoccoz 1991; Kirk 1996; Robinson and Wainer 2002). In reality, two populations will rarely ever be exactly the same, but will vary to differing degrees. With control charts, one is able to see how far the observations fall from both the centerline and the established control limits. Additionally, with control charts, one would not necessarily have to conclude that no difference exists over time in the parameter of interest, but rather that the observations are either within the specified range of acceptable values or not.

One advantage of control charts compared to a traditional NHST approach is that a number of potential nonrandom patterns may be evident from a control chart (as described above) and each type of pattern may contribute a unique component of information regarding the indicator. With a NHST approach, the primary conclusion is simply to reject or fail to reject the null hypothesis. Although the tests associated with NHST may provide additional information, there is frequently a one-dimensional focus on the *P*-value.

Perhaps the primary advantage of using a NHST approach in a monitoring program is the fact that such an approach necessitates a yes-no decision (i.e., either the parameter of interest is changing or it is not). In such a case, rejection of the null hypothesis may trigger management action. The primary advantage of using control charts over a pure parameter estimation approach is that control charts establish a threshold for action, whereas simple parameter estimation with associated confidence intervals does not.

In environmental monitoring, a frequent goal is to determine when an indicator has changed by a certain percentage (e.g., when a population has declined by 20%). Such a change is presumed to represent a biologically meaningful level at which management should be concerned. Yet, this is not easy to accomplish with a traditional NHST approach. A power analysis is frequently employed in this context, yet one must be careful not to confuse the 'effect size' of the power analysis with the amount of absolute change in the population (Morrison 2007). In other words, with a traditional NHST approach, one is testing whether there is some non-zero difference, and thus any conclusions must relate to the hypotheses under test. For any test, the effect size for a given level of power will vary depending upon the other variables that enter into the calculation of power, such as sample size and the probability of a Type I error. Testing biologically meaningful hypotheses (e.g., a null hypothesis of <20% change) requires the use of a different type of sampling distribution than is available in most texts and computer software packages, and such null hypotheses are much more difficult to reject (Murphy and Myors 2004). In contrast, control charts allow one to easily determine when an indicator has changed by any given percentage. All one has to do is establish threshold limits as a percentage of the baseline value.

A HYPOTHETICAL EXAMPLE

Consider the following example: A population of concern is monitored on an annual basis. Each year, five surveys of the population are conducted on separate days over a two-week period. (We assume the population consists of long-lived organisms and will change negligibly over a two-week period.) Nine years of data are currently available. Management would be concerned if the population was found to decrease by 20%, and in that case would implement measures designed to further protect the population.

Three simple analysis techniques are applied to the data: (1) linear regression, (2) parameter estimation with confidence intervals, and (3) a control chart approach. We assume that the data come from a normal distribution, which is required by the first two techniques, but not the control chart application in this particular example.

The regression is an inferential approach that tests the hypothesis that the population is changing (in this case, linearly) over time. In the example, the sample mean estimates are regressed against time (Figure 3A). The regression reveals no significant trend by the traditional $\alpha = 0.05$ threshold ($y = -2.06\ x + 105.31$, $R = 0.623$, $P = 0.073$). If one assumes the population has only recently begun to decline (see below) and uses only the last four years in the regression, then $P = 0.33$, due to the small sample size.

Parameter estimation with associated confidence intervals is primarily a descriptive approach that, particularly in this example, conveys information on the uncertainty surrounding our estimates. Because our sample size for each point is relatively small (n = 5), the confidence intervals are relatively wide (Figure 3B). (The sample size chosen, however, may not be unrealistic for many environmental monitoring efforts.) The potential problem with this approach, as mentioned above, is in determining which points in time are "significantly" different from other points. For example, in comparing two independent means, 95% confidence intervals may overlap to a certain extent and yet the means would still be considered statistically different at $P = 0.05$ (Schenker and Gentleman 2001; Cumming and Finch 2005).

For the control chart approach, we plot the sample means for each year. The first five years are selected to represent a baseline, which, in this example, represents a population size of 100. This selection is

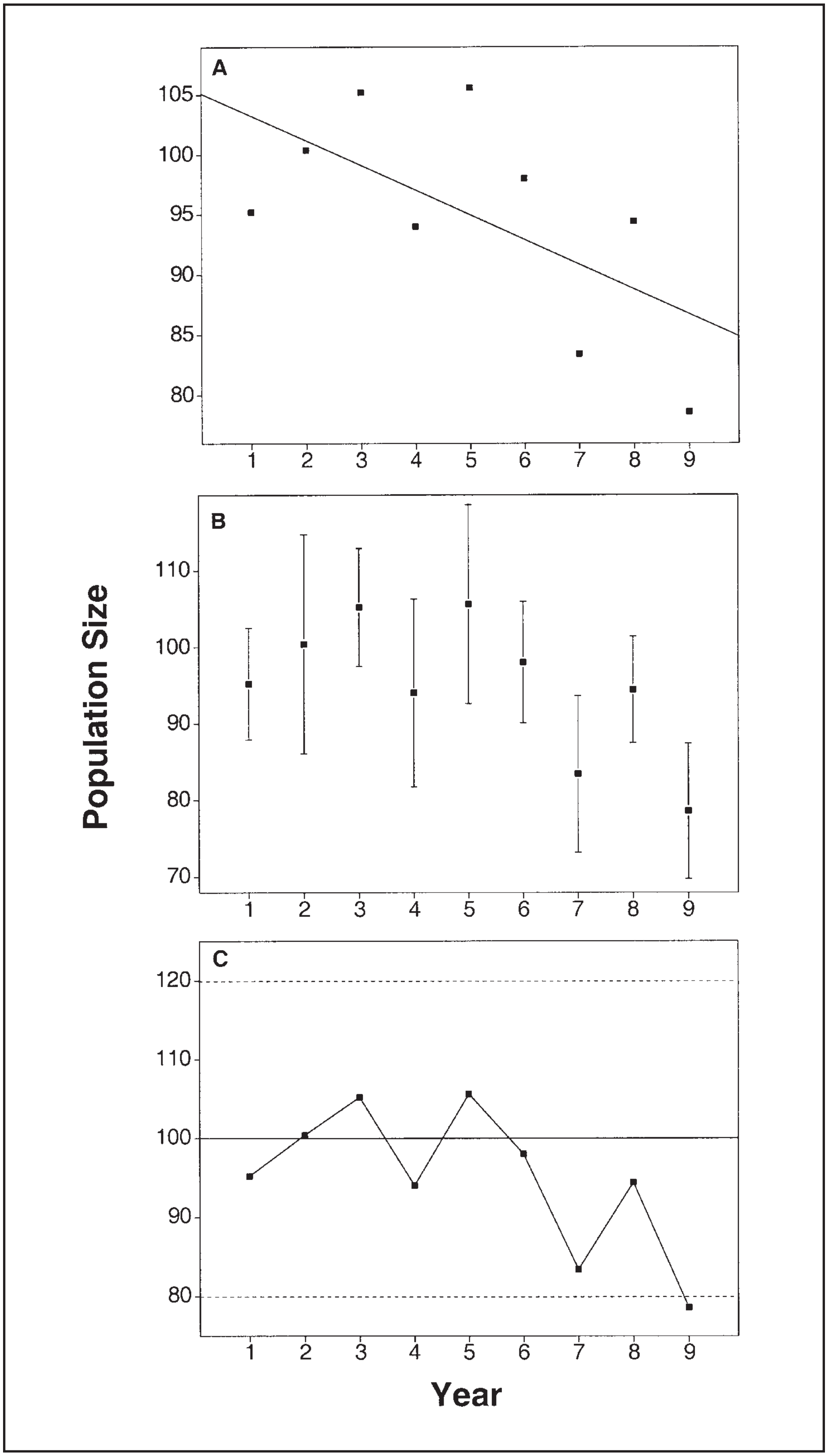

Figure 3. Analysis of the same hypothetical data set by (A) linear regression, (B) parameter estimation with 95% confidence intervals, and (C) a control chart. The solid line in C represents a baseline determined from the first five years; the dashed lines represent control limits that indicate a 20% change from this baseline.

somewhat arbitrary, because the first five (or six) years seem to vary randomly about a stable mean. Because of management's voiced concern of detecting a 20% change, we set a control limit at 80 (and also at 120, to be able to detect meaningful changes in either direction). In year nine, the population does decrease below the 20% change threshold, and triggers an alarm for management (Figure 3C).

Admittedly, one can contrive an example to support any argument, and in many cases, control charts may not represent a clearly advantageous approach. Another year or two of data and the regression in the example may become significant at $\alpha = 0.05$. Yet, if we are dealing with an endangered population in decline, the additional delay could prove costly. Of course, a less rigid view of interpreting *P*-values would help in many cases employing hypothesis tests.

Perhaps the biggest advantage of parameter estimation with confidence intervals is that the amount of variability in the parameter of interest is obvious, whereas with traditional control charts, it is hidden. Of course, there is no reason why confidence intervals, or some other measure of variability around the means (e.g., standard errors), could not be portrayed in control charts.

EXCEPTIONS

Although control charts could be employed in the majority of natural resource monitoring efforts, they will not be effective in all cases. For example, control charts are not likely to be useful in the case of populations that exhibit extreme temporal variation (e.g., Thomas 1996), as it would be difficult to establish a meaningful baseline. Large amounts of variability, however, will be problematic for any attempt to demonstrate a significant trend (Gibbs et al. 1998). If the variable of interest is changing from the inception of the monitoring program, so that it is not possible to establish a stable baseline, a regression analysis may be more informative. Data sets consisting of presence-absence or categorical data may not easily lend themselves to the use of control charts. (These types of data preclude many other analyses as well.) Control charts could still be employed, however, to a certain extent, analogous to the analysis of attribute or count data in industry. For example, one could focus on the percentage of sites containing a certain species or on the number of plots matching a certain category of interest.

CONCLUSIONS

Control charts hold many potential advantages for management: they are relatively easily to construct, their interpretation is straightforward, they provide a standard format for evaluation of many different variables, they can be easily updated whenever additional data become available, the relative risk of Type I and Type II error are intuitive and easily adjusted, and one may define a threshold for action at any desired level. They can be constructed for many different types of indicators – whether univariate or multivariate. Control charts may reveal a diversity of different types of non-random patterns, which may indicate various potential concerns with the resource in question. For a given distribution, one may calculate a number of post hoc probabilities, most of which are simple and could be done by almost anyone with a basic knowledge of mathematics. Ultimately, all one really needs to know to interpret a control chart correctly is what the centerline represents, what the control limits represent, and what sort of distribution (if any) is assumed.

The primary disadvantage of the control chart approach has to do with the difficulties of determining the 'normal' or baseline state for the variable in question. Establishing the appropriate baseline will likely require an adequate amount of good judgment in addition to empirical data. A limitation of control charts is that they do not indicate a cause for any changes in the indicator of interest. Yet, few approaches to monitoring trends do, and manipulative experiments are usually required to determine cause and effect. Control charts may be used, however, to evaluate correlations with known disturbance events (as in Anderson and Thompson 2004).

Finally, it should be noted that control charts do not preclude the use of other more detailed or sophisticated statistical analyses. Rather, control charts represent a potential basic analysis for almost any data set – a sort of 'quick look' for busy managers to determine which variables are in the greatest need of more in-depth analyses.

ACKNOWLEDGMENTS

This paper benefited from helpful discussions with Mike DeBacker.

Lloyd Morrison is a Quantitative Ecologist with the Heartland Network of the National Park Service's Vital Signs Inventory and Monitoring Program and Adjunct Professor at Missouri State University.

LITERATURE CITED

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.

Anderson, M.J., and A.A. Thompson. 2004. Multivariate control charts for ecological and environmental monitoring. Ecological Applications 14:1921-1935.

Atkinson, A.J., R.N. Fisher, C.J. Rochester, and C.W. Brown. 2003. Sampling design optimization and establishment of baselines for herptofauna arrays at the Point Loma Ecological Reserve. U.S. Geological Survey, Western Ecological Research Center, Sacramento, Calif.

Beauregard, M.R., R.J. Mikulak, and B.A. Olson. 1992. A Practical Guide to Statistical Quality Improvement: Opening Up the Statistical Toolbox. Van Nostrand Reinhold, New York.

Belia, S., F. Fidler, J. Williams, and G. Cumming. 2005. Researchers misunderstand confidence intervals and standard error bars. Psychological Methods 10:389-396.

Carver, R.P. 1993. The case against statistical significance testing, revisited. Journal of Experimental Education 61:287-292.

Cherry, S. 1998. Statistical tests in publications of The Wildlife Society. Wildlife Society Bulletin 26:947-953.

Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117-143.

Colegrave, N., and G.D. Ruxton. 2003. Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. Behavioral Ecology 14:446-450.

Conover, W.J. 1999. Practical Nonparametric Statistic, 3rd ed. J. Wiley, New York.

Cumming, G., and S. Finch. 2005. Inference by eye: confidence intervals and how to read pictures of data. American Psychologist 60:170-180.

Cumming, G., J. Williams, and F. Fidler. 2004. Replication and researchers' understanding of confidence intervals and standard error bars. Understanding Statistics 3:299-311.

Dixon, P.M., A.R. Olsen, and B.M. Kahn. 1998. Measuring trends in ecological resources. Ecological Applications 8:225-227.

Elzinga, C.L., D.W. Salzer, J.W. Willoughby, and J.P. Gibbs. 2001. Monitoring Plant and Animal Populations. Blackwell Science, Osney Mead, Oxford, U.K.

Gaston, K., and B.H. McArdle. 1994. The temporal variability of animal abundances: measures, methods, and patterns. Philosophical Transactions of the Royal Society of London, Series B 345:335-358.

Gibbs, J.P., S. Droege, and P. Eagle. 1998. Monitoring populations of plants and animals. BioScience 48:935-940.

Gibbs, J.P., and S.M. Melvin. 1997. Power to detect trends in waterbird abundance with call-response surveys. Journal of Wildlife Management 61:1262-1267.

Gyrna, F.M. 2001. Quality Planning and Analysis: From Product Development Through Use, 4th ed. McGraw-Hill Irwin, New York.

Hatfield, J.S., W.R. Gould IV, B.A. Hoover, M.R. Fuller, and E.L. Lindquist. 1996. Detecting trends in raptor counts: power and Type I error rates of various statistical tests. Wildlife Society Bulletin 24:505-515.

Hayes, J.P., and R.J. Steidl. 1997. Statistical power analysis and amphibian population trends. Conservation Biology 11:273-275.

Hoenig, J.M., and D.M. Heisey. 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician 55:19-24.

Johnson, D.H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763-772.

Kirk, R.E. 1996. Practical significance: a concept whose time has come. Educational and Psychological Measurement 56:746-759.

Lindley, S.T., M.S. Mohr, and M.H. Prager. 2000. Monitoring protocol for Sacramento River winter Chinook salmon, *Oncorhynchus tshawytscha*: application of statistical power analysis to recovery of an endangered species. Fishery Bulletin 98:759-766.

Manly, B.F.J. 2001. Statistics for Environmental Science and Management. Chapman & Hall/CRC, Boca Raton, Fla.

Mardia, K.V. 1971. The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. Biometrika 58:105-121.

McBean, E.A., and F.A. Rovers. 1998. Statistical Procedures for Analysis of Environmental Monitoring Data and Risk Assessment. Prentice Hall PTR, Upper Saddle River, N.J.

Montgomery, D.C. 2001. Introduction to Statistical Quality Control, 5th ed. J. Wiley, New York.

Morrison, L.W. 2007. Assessing the reliability of ecological monitoring data: power analysis and alternative approaches. Natural Areas Journal 27:83-91.

Murphy, K.R., and B. Myors. 2004. Statistical Power Analysis: a Simple and General Model for Traditional and Modern Hypothesis Tests, 2nd ed. Lawrence Erlbaum Associates, Mahwah, N.J.

Nakagawa, S. and T.M. Foster. 2004. The case against retrospective power analyses with an introduction to power analysis. Acta Ethologica 7:103-108.

Olson, C.L. 1974. Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statistical Association 69:894-908.

Pettersson, M. 1998. Monitoring a freshwater fish population: statistical surveillance of biodiversity. Environmetrics 9:139-150.

Robinson, D.H., and H. Wainer. 2002. On the past and future of null hypothesis significance testing. Journal of Wildlife Management 66:263-271.

Schenker, N., and J.F. Gentleman. 2001. On judging the significance of differences by examining the overlap between confidence intervals. The American Statistician 55:182-186.

Shewhart, W.A. 1931. Economic Control of Quality Control. Reinhold, Princeton, N.J.

Taylor, L.R. 1961. Aggregation, variance, and the mean. Nature 189:732-735.

Thomas, L. 1996. Monitoring long-term population change: why are there so many analysis methods? Ecology 77:49-58.

Utts, J. 1988. Successful replication verses statistical significance. The Journal of Parapsychology 52:305-320.

Van Strien, A.J., R. Van De Pavert, D. Moss, T., Yates, C.A.M. Van Swaay, and P. Vos. 1997. The statistical power of two butterfly monitoring schemes to detect trends. Journal of Applied Ecology 34:817-828.

Wetherill, G.B., and D.W. Brown. 1991. Statistical Process Control: Theory and Practice. Chapman & Hall, London.

Yoccoz, N.G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America 72:106-111.