POWER ANALYSIS FOR LONG-TERM MONITORING OF FISHES IN SELECTED WATERS OF THE GILA RIVER BASIN, ARIZONA

June 2000

Submitted by:

Linda J. Allison
Nongame and Endangered Wildlife Program
Arizona Game and Fish Department
2221 West Greenway Road
Phoenix, Arizona 85023

SUBMITTED TO:

U.S. Bureau of Reclamation
Phoenix Area Office
P.O. Box 81169
Phoenix, Arizona 85069-1169

Federal funding provided under
Cooperative Agreement No. 99-FG-32-0200

TABLE OF CONTENTS

FIGURES

TABLES

APPENDICES

POWER ANALYSIS FOR LONG-TERM MONITORING OF FISHES IN SELECTED WATERS OF THE GILA RIVER BASIN, ARIZONA

Linda J. Allison

INTRODUCTION

IMPETUS FOR THE MONITORING PLAN

Completion of the Central Arizona Project (CAP) in 1993 allows the potential transport of non-indigenous fishes and other aquatic organisms from the Colorado River, where the CAP network of canals originates, to central and southern Arizona. A U.S. Fish and Wildlife Service (USFWS) Biological Opinion (BO) on transportation and delivery of CAP water to the Gila River Basin (USFWS 1994) determined that the project would jeopardize continued existence of 4 threatened or endangered fishes: Gila topminnow (*Poeciliopsis occidentalis*), spikedace (*Meda fulgida*), loach minnow (*Tiaroga cobitis*), and razorback sucker (*Xyrauchen texanus*).

Reasonable and Prudent Alternative No. 2 (USFWS 1994) directed the Bureau of Reclamation (USBR) to develop a monitoring plan in conjunction with USFWS and Arizona Game and Fish Department (AGFD). USBR was directed to determine baseline community composition and distribution and to monitor impact of these non-native fishes on existing fish communities in the Gila and San Pedro Rivers. In addition, movement of non-natives from the CAP canals was to be described.

POWER ANALYSIS

History and the current project

The BO calls for monitoring in a specific set of areas: 1) the CAP aqueduct; 2) Salt River Project (SRP) canals; 3) Florence-Casa Grande (FCG) Canal; 4) Salt River between Stewart Mountain Dam and Granite Reef Dam; 5) Gila River between Coolidge and Ashurst-Hayden dams; and 6) perennial reaches of the San Pedro River downstream from the U.S.-Mexico border. The BO requires that baseline data and trend detection be performed for each of these areas.

Using preliminary data, a power analysis (Wilson 1996) reported on the ability of the USBR monitoring plan (Clarkson 1996) to detect changes in community composition. Since then, sampling methods have been standardized and a second power analysis would be appropriate. Specifically, AGFD reports here on viability of the monitoring plan for detecting changes in current community structure on the Gila and San Pedro Rivers (Figure 1).

To evaluate suitability of the current spatial scale for describing abundance and distribution, the analysis first describes spatial and temporal consistency of the fish assemblage. The power analysis also involves estimation of within-site variability (standard deviation and coefficient of variation) for each species. These variability estimates are used in a power analysis consistent with the model to be used in final analyses of data.

The draft monitoring protocol specifically calls for examination of the tradeoff between increasing levels of effort and levels of change we want to detect, so this report examines the effects of changing sample length and number as well as number of years to detect a change.

The monitoring protocol makes provision for quantitatively describing and detecting trends in the abundance of more common species, as well as for expending sufficient effort to detect uncommon species. There is no operational definition of "common," of course, and USBR's interest in tracking abundance of some less common species might blur the distinctions between how common and uncommon species are treated. This report explores ability of the monitoring protocol to 1) describe trends, and 2) build density estimates, or 3) simply detect each species.

What is a power analysis?

A common goal in monitoring projects is to determine whether population size or diversity is increasing or decreasing. A linear regression of estimated abundance (or an index of diversity) against time is commonly used to evaluate such a trend (Gerrodete 1987, Gibbs and others 1998). By definition, a trend is detected when the regression has a slope significantly different from zero. The conclusion that a trend in abundance is occurring, when in fact it is not, is termed a Type I error (α), while the conclusion that no trend is occurring, when in fact it is, is a Type II error (β). Power is the probability of detecting a trend when it is occurring ($1 - \beta$). Ability to distinguish a trend from an underlying baseline condition of spatial and temporal variability will depend on 1) the amount of underlying variability, 2) the strength of the trend (the effect size), 3) the number of samples taken at each point in time and space, and 4) the number of years over which the monitoring occurs, because the longer any trend impacts the population, the more the final count will differ from the initial count.

Because many factors interact to determine the statistical power of a monitoring program, power estimation is a complex process. In addition to specifying an effect size of interest, and considering possible study designs, there is the necessity to deal with properties of the index itself (Gibbs and others 1998). For instance, as counts increase or decrease over time, is this expected to have an effect on variance? If so, how can we account for this effect on our ability to detect a trend? At a more fundamental level, what statistical model

will be used to identify a trend in community structure? When we are interested in effects on more than one species, our description of change, indeed, whether we identify change at all, will be a function of the number and type of species we use as our statistical definition of an assemblage (Yant and others 1984, Herbold 1984, Rahel and others 1984)

The focus of this paper is on a specific part of the study design; namely, what modifications of the proposed sampling design can improve ability to detect changes in the chosen index (counts per unit stream length). That the index is a fair surrogate for actual animal densities is an assumption of this paper. I merely note, as does the monitoring protocol, that this index can be problematic, especially when one method must suffice to describe a suite of species, all with different vulnerabilities to the capture method. That vulnerability will change under different habitat conditions. Clarkson (1996) addresses some of these concerns and how one might nonetheless make the best estimate of species densities. Variance components used for the current analysis reflect variability ascribable (but not necessarily identifiable) to the sampling method (Gibbs and others 1998). Simply put, changes in the protocol will change variance and therefore power of the analysis. Some positive changes are considered here. Others, such as the capture method and experience level of the field crew are not considered here.

DESCRIBING TRENDS IN ASSEMBLAGES

Evaluating changes in absolute abundance, Rahel (1990) pointed out that we can describe assemblages by 1) the absolute count or density of individuals in each species, 2) the abundance ranking of each species, or 3) the presence of each species. These descriptions form a nested set: If an assemblage is stable when described in terms of density of each species, the abundance rankings and presence/absence structure will also be stable. The reverse is not true, because assemblages that have very consistent species composition (presence/absence) may be characterized by unstable abundance rankings and/or abundances. Rahel (1990) also noted it is most difficult to detect changes in absolute abundance, but easier to detect changes in relative frequencies or presence/absence structure. In other words, exhaustive sampling is required to detect changes in abundance in the face of spatial and temporal variability, whereas changes in relative abundance seem detectable using lower effort methods, such as one-pass sampling (Simonson and Lyon 1995).

In this analysis, the fundamental question to be addressed by results of the long-term monitoring is whether fish associations are changing. However, the BO stipulates specifically that abundance of individual species will be monitored. For more common species, we may be able to describe whether their abundance is changing in individual reaches. For uncommon species, where our most solid information concerns whether the

species is present (it is more difficult to give accurate density estimates), we might answer the same question by testing whether there has been a change in the number of reaches where the species was encountered in a given river (Strayer 1999). Because USBR already stipulates monitoring all accessible reaches with perennial water (Clarkson 1996), the number of reaches available to describe occupied proportion is limiting.

The precision of a point estimate for relative frequency is entirely related to the number of individuals sampled. This means that unless the same number of individuals in total was sampled each time, comparisons of different relative frequencies are based on a different degree of precision for each estimate. This problem is compounded because detection ability is sensitive to number of fish sampled, so different sized samples also have different likelihood of detecting uncommon fish. Especially for rare species, these proportions are very sensitive to the total number of fish in the sample (Green 1979). This idea leads to the alternate suggestion to sample approximately the same number of individual fish at each station, instead of sampling the same number of linear meters (Peet 1975, Angermeier and Schlosser 1989, Wilson 1996). The current protocol uses fixed-length samples, and there is considerable variability between 200 m stations in the total number of fish sampled (N=0 to 5933), although the majority are within an order of magnitude of one another.

METHODS

DESCRIPTION OF THE SAMPLING REGIME

The current analysis uses data collected on fish in the Gila and San Pedro Rivers (Table 1). Although USBR also visited these sites from 1991 to 1994, the current protocol (Clarkson 1996) has only been implemented since 1995. In this report, the phrase "baseline data" refers to data collected annually from 1995 through 1998. The 7 reaches indicated in Figure 1 were chosen within the area covered by the BO, using geomorphic criteria such as channel and floodplain width (Clarkson 1996). Stations sampled in each reach were chosen for accessibility and presence of perennial water.

The protocol calls for single-pass "quantitative" sampling at each of three 200 m stations at each reach. Fish are identified to species and age class (first-year or older). Within each 200 m station, macrohabitats are first sampled in an upstream progression using single-pass electroshocking. Macrohabitats with slow-moving waters are sampled in an upstream direction; fast-moving waters may be sampled in a downstream direction.

In order to detect uncommon species, Lyons (1992) recommended sampling a distance along the length of the stream equal to 35 times the stream width; 200 meters represents

35 times the width of a stream that is 5.7m wide. However, Paller (1995) recommends a considerably longer station length for single-pass sampling, 100 to 450 times the stream width. This uncertainty in adequate sample lengths led USBR to propose supplemental sampling to enhance ability to detect invading species. Species abundances in different stations are to be compared quantitatively only if the fish were collected using an electroshocker. However, species lists should be supplemented by using other gears after all electrofishing is completed. Uncommon habitats in contiguous areas should also be included in these "qualitative" samples.

CONCORDANCE AND CORRELATION BETWEEN BASELINE SAMPLES.

Because baseline data have been collected, before proceeding with the power analysis itself, this report examines spatial patterns that currently exist. This helps identify potential changes that might occur in the assemblage. Preliminary examination of patterns will point to statistical assumptions that may be violated and that will have to be addressed in the future. This analysis of baseline data includes simple examination of spatial distribution of each species. Do samples taken outside the standard protocol change our view of this spatial pattern?

The report continues by examining the degree to which samples vary spatially and temporally. To investigate the consistency of the assemblage between years, and concordance between reaches sampled in the same river in the same year, $\chi 2$ analysis was used to test for concordance of absolute counts of species (Grossman and others 1982, Moyle and Vondracek 1985). Kendall's W was used to test for constancy of abundance rankings. Friedman's (Siegel and Castellan 1988) would also be appropriate, but Kendall's W is standard in community ecology literature (Grossman and others 1982, Moyle and Vondracek 1985, Rahel 1990, Kendall and others 1992), and in this case the two tests will give identical results.

Finally, for each species, all pairs of counts taken in the same reaches in sequential years are used to test for autocorrelation. If temporal autocorrelation exists, slope estimates will not be biased, but the estimates will vary considerably more than indicated by error estimates, which are underestimated. The effect of autocorrelation on ability to detect trends is based on the fact that 1) sample trend estimates are highly variable (a function of the initial sample error in the first count taken), and 2) error estimates are too small. The effect will be to increase likelihood of Type I and Type II errors.

MODEL USED FOR THE POWER ANALYSIS

A main-effects model without autocorrelation.

Because stations cannot always be sampled (due to drying, for instance), and because the protocol is supposed to detect changes at each reach, stations were treated as sampling units. Reaches are treated as fixed effects in an analysis of covariance (ANCOVA), with time (measured in years) as a covariate. Under the most straightforward scenario, any trends will be the same at all reaches, which means there will be no interaction between time and reach effects. Indeed, because power analyses compare the null hypothesis to a specific alternative, such analyses are usually limited to consideration of main-effects models. In this particular case, there is suspicion that not all reaches will reflect the same trends in species abundances, should an impact occur. For instance, reaches closest to the CAP might show more dramatic change earlier. However, there has been no formal attempt to predict how fast any impacts from the CAP canal might spread, and it is easiest to consider specified main effects, but difficult to cover the set of possible interaction effects and magnitudes. Therefore, the model addresses whether a system-wide change of uniform magnitude can be detected.

Treating time as a covariate is the simplest approach to modeling change over time. It allows us to consider a power analysis that looks for a particular effect size (for instance, a 20% decline). Under the alternative, a repeated measures scenario, there are many ways for the beginning and ending time point to reflect a similar proportional change, but the intervening change would not necessarily be linear, nor can we stipulate *a priori* what shape that change might take. Recent work suggests that descriptions of change using linear regression perform well for power analysis (Harris 1986, Gerrodette 1987, Eberhardt and Simmons 1992, Gibbs and others 1998,). It is also possible to consider reaches as repeated measures within each year, but this was not pursued because there is a sense that reaches might not change in parallel (Gurevitch and Chester 1986), so the final analyses performed by USBR will need to test this possibility. By treating time as a covariate, it is possible to examine interactions between the effects of time and of reach.

Use of Monte Carlo simulations

Basic analytical means of evaluating power were considered initially (Gerrodette 1987, van Strien and others 1997, Gryska and others 1997). Another option, however, is to use Monte Carlo simulations. This type of randomization analysis (Manly 1997) simulates thousands of different monitoring outcomes given variance and mean count estimates from our baseline data. This gives us the possible outcomes if no trend exists. We can also incorporate potential trends, or the effect of sampling more stations at each reach. By comparing the outcomes from these simulations, we can evaluate how robust our ability is, using a given protocol, to see important trends. Peterman (1990) refers to Monte Carlo simulation as a means of generating power analyses when analytical formulae for power have not been derived, such as in complex sampling designs (Peterman and Routledge 1983, Peterman and Bradford 1987).

Distribution used to describe counts

Three basic distributions are used to model count data: the binomial, negative binomial, and Poisson distributions. The binomial distribution is easiest to apply when the counts for one species make up a large proportion of the sample. If the species occurs less frequently, less than 5% of the sample, then the Poisson distribution can be used to approximate the binomial. If the species is uncommon and counts per sample are of interest, then the Poisson is appropriate on its own. The negative binomial is used when individuals do not occur at random, so that samples reflect different degrees of clumping.

A fourth distribution, the lognormal, is not strictly applicable to count data. The lognormal distribution is a continuous distribution, although like the above discrete distributions, it is bounded below by zero. (Count data cannot be appropriately modeled by distributions with negative values.) Another desirable property of the lognormal distribution is that it can simultaneously model species that occur at low and high densities. On the other hand, to use the Poisson distribution for simulations for common and uncommon species, each species would have to be modeled at a different spatial scale so that average counts do not exceed a small number, say 5. Finally, percent change is implied if log-transformed data are used. (Green 1989, van Strien et al 1997). That is, whereas a change of 0.2 units in absolute counts may be either a large or small change, depending on whether initial counts are low or high, a change of 0.2 units on a log scale always reflects a 20% change in abundance. This property is useful when a regression analysis is used, so that a slope of 0.2 describes a 20% change in abundance for one unit of time.

PROGRAM MONITOR

Input to the program

This program (Gibbs 1995; software available at URL: http://WWW.MP1-PWRC.USGS.GOV/powcase/monitor.html) was used for the power analysis because it has many helpful properties, including availability for rapid testing of other scenarios, should USBR become interested in other protocol options. The program uses Monte Carlo simulations to generate counts from a lognormal distribution, rounding all count estimates to the next-lowest integer. In order to allow for zero counts (for which no log value exists), 1 was added to all counts.

Input for the program allows us to stipulate the number of reaches we are considering (plots) and number of stations per reach (counts/plot/survey). Simulations also considered the effect of changing length of station (plot counts); by doubling the baseline counts, stations are simulated that are twice as long. The average counts at each reach allow the program to account for between-reach variability. Temporal variability, estimated for each species from baseline data, are the final input to the simulation. For

each simulation, the program generates power estimates to detect annual trends ranging from –10 to +10 %. The same annual trend, acting over a longer period of time, will have a larger impact than over a shorter time frame, so increasing the length of the reporting period can affect ability to detect trends (Table 2). Simulations run for each species are listed in Table 3.

Although several modifications to the protocol were simulated, it was assumed that the number of reaches would not change (all accessible perennial reaches provide important coverage). Also, given the different flow patterns of each river and canal that is sampled (Clarkson 1996), sampling more stations on one monitoring expedition may be logistically or financially simpler than to coordinate smaller excursions to fewer stations more frequently during the year, every year. For this reason, it was assumed that sampling more than once a year was not practical. The BO does not specify the level of change (effect size) that might be of concern. For this reason, I present results for the range of annual trends simulated by this program (–10% to +10%).

Risk of Type I error

When managing small numbers of individuals in rare species, the consequences of Type I error (incorrectly concluding that a population has declined) are less severe than failing to detect a decline when one has occurred (Type II error). On the other hand when managers are looking for an increase in a common and potentially troublesome species, the risk of a Type I error may be more costly than a Type II error. This analysis addresses species fitting both scenarios. For this reason, I simulated α levels of 0.05, 0.10, and 0.20, and used 2-tailed tests (because the project will be considering both increases and decreases, should they occur). Gibbs and others (1998) argue that for purposes of monitoring we should consider conventional values of α=0.05 and β=0.20. Because detection of a trend when it is occurring is a goal, we would usually not accept study designs with power below 0.8. Other authors have discussed relative risk, pointing out that if we set α=0.05 and β=0.20, then we are accepting a Type II error rate that is 4 times higher than our Type I error rate. This perspective has led to the suggestion that we begin with equal Type I and II error rates, adjusting them according to relative consequences of each type of error (Cohen 1988).

Estimating temporal variability

In addition to mean counts at each reach, the simulations require an estimate of the degree to which counts at each reach vary year-to-year when no trend is present (temporal variability). For these estimates, I used baseline data fit to an ANOVA model with reach as the main effect. Residual variation after fitting the model is described by a variance term, the mean-squared error (MSE). For Program Monitor, variability is described using the standard deviation as fraction of the mean counts (the coefficient of variation, CV), so the square root of MSE (RMSE) was taken to approximate the

standard deviation of counts within each reach (Gibbs and others 1998). The RMSE of log-transformed data is the CV itself. Although variance usually increases with the mean, this is not expected when the data are log-transformed.

DETECTING CHANGE IN RELATIVE FREQUENCY

Power analyses to examine different sampling designs were only conducted using absolute counts. However, for a particular sampling design, a different type of power analysis can be used to examine the amount of variability that gives acceptable power. Variance in a proportion is a function of the sample size (for instance, number of fish of all species) and of the proportion that is estimated, so I generated graphs to illustrate this dependency.

RESULTS

PATTERNS OF CONCORDANCE AND CORRELATION

Table 4 illustrates that not all sampling stations were sampled each year. The San Pedro stations were sampled consistently using quantitative sampling, but the analysis proceeded by assuming that only reaches will be consistently sampled each year, while stations are drawn as replicates from each reach. There is no assumption that the same stations are sampled each year.

Following Moyle and Vondracek (1985) in their definition of species assemblage, Table 5 lists 1) species found at all reaches in each river during at least 1 year of sampling, and 2) species found at least once in all reaches on a particular river. Note that the Gila and San Pedro are not characterized by the same species assemblage (also see next section). Using α=0.05, absolute counts were not consistent between years within reaches (Table 6) or between reaches within years (Table 7). However, rank abundances did show patterns of consistency. Reaches on the San Pedro all had between-year concordance, but only in 1995 did the San Pedro show reach-to-reach agreement. On the Gila River, in contrast, year-to-year concordance was only the norm for 1 of the 4 reaches (reach 7); however, within years, rank abundance of species was in agreement between reaches.

Serial autocorrelation was detected by considering the absolute frequency of each species for each pair of years within reaches (Table 8). Species had as few as 1 year of observations (MIDO) to 15 pairs of observations (AGCH, AMNA, GAAF, LECY, PACL). Although patterns at each reach are of interest because the monitoring plan must describe the spatial component of any impacts, the presence of serial correlation implies that we would have to account for the effects of each reach anyway. Because the degree of serial autocorrelation differs by species, autocorrelation was not incorporated into the model. For species with serial autocorrelation, due to the resulting underestimate of variability, results of the power analysis will reflect more power than should actually be expected.

SPECIES DISTRIBUTION, TEMPORAL VARIANCE, AND BASELINE COUNTS FOR EACH REACH

Table 9 reports CV and mean counts used for simulations for each species. Values for CVs are within the range of those reported for other studies and groups of organisms (Gibbs and others 1998). Table 9 also illustrates the patterns of distribution and abundance that are seen in the current species assemblage. Some species have been detected in every reach that is currently monitored (AGCH, AMNA, GAAF, LECY, PACL). Other species have only been detected a few times and are associated with Gila reaches near San Carlos Reservoir (MIDO, PONI, PYOL). Another group of species is found in the Gila reaches and on the San Pedro River at the convergence with the Gila River (CAIN, CYCA, CYLU, DOPE, ICPU). MISA is also found on the Gila River, but has been detected at the uppermost reach on the San Pedro, not at the confluence. Finally, one set of species is found primarily on the San Pedro River (LEMA, AMME, PIPR), although the first species is quite uncommon.

ABILITY TO DETECT TRENDS WITHIN SPECIES

Power using the current monitoring protocol

Simulation results are given in Appendix 2. Figures 2 through 5 illustrate power of the current monitoring protocol to detect different trends over a 10 year period. In the following discussions, power of 80% or greater is considered to be sufficient. Each figure portrays power for species with similar spatial distributions, as discussed above. Note that as long as all reaches reflect the same overall trend (an assumption of this analysis), the current monitoring protocol has sufficient power to detect changes in species that are found at more reaches (Figure 2). This pattern may reflect the fact that species found in fewer reaches are also at lower densities, so trends will be more difficult to detect (Figure 6). As Appendix 1 indicates, some species (DOPE, LEMA, MIDO, PONI, PYOL) have been detected only a few times in so few reaches that no alterations of the current protocol will allow for detection of the annual trends considered here. Neither these species, nor the 4 threatened or endangered species, which have not been detected during baseline monitoring, will be considered further in discussion of detection of trends in abundance. Trends of less than 10% increase or decrease in AMNE, MISA, PIPR cannot be detected with reasonable power using the current protocol and standard Type I error rates.

For species that occur in many reaches and for which we can detect meaningful trends, the next question is whether trends in individual reaches can be detected over the same time frame. Comparison of Figure 2 to Figures 7 and 8 illustrates that only if a species occurs initially at high densities can a trend be detected in a single reach. Where they currently occur, only CAIN and CYLU exist at high enough densities for modeled changes to be detected in each reach. In summary, only relatively abundant species can be considered for trend detection using the current protocol. “Abundant” in this context

may refer to either density within a single reach or to occurrence in several reaches, even at low densities.

Power under alternate scenarios

Effort can be modified by changing the number of stations sampled per reach or by changing reach length. Figure 9 illustrates effect of different types and levels of changing effort. In general, a single station per reach performs poorly, with increasing power as number of stations increases; however, there does not seem to be a clear advantage to sampling 4 stations instead of 3. Increasing station number from 3 to 4 did not have much effect on species that have either very low or very high power for monitoring initially. For instance, although there is a clear pattern of increasing station number also increasing power to detect changes in MISA, the increase is not sufficient to provide adequate power (Figure 10). Increasing station length also improves power, and in cases where power is insensitive to number of stations, increasing station length can have a more dramatic effect (Figure 10). To directly compare alternative means of increasing power by increasing effort, note that sampling two 400-m stations improves power considerably, although sampling four 200-m stations does not. A similar pattern was observed for ICPU, LECY, and PIPR, and trivial improvement was seen for a few other species.

The effect of decreasing station number was also simulated. Power to detect trends for most species was unaffected by whether 3 or 2 stations are sampled. Obviously, species that perform poorly under the current protocol will still perform poorly if fewer stations are sampled. However, a few species that perform adequately under the current protocol would not perform adequately by eliminating even one station from sampling (AGCH, AMNA, ICPU, LECY; Appendix 2).

Power to detect trends for some species would be improved to over 80% by using α=0.10 instead of α=0.05. The species for which power would be increased (MISA, LECY, CYCA, ICPU) all show borderline performance under the current protocol, so adopting a different α will depend on how biologists weigh the risks of Type I and Type II error. In the case of MISA, only α of 0.20 brought the analysis up to reasonable standards (Figure 11).

On average, it is easier to detect the same annual trend after this trend has impacted a population for a longer time span than a shorter one (Table 2, Figure 12). However, this is not true when discussing declining trends in species that are currently uncommon. For instance, for MISA and PIPR, some positive trends could probably be detected with high power after 20 years of monitoring, but negative trends will not be detectable. Even after 20 years, significant trends in AMME would not be detectable (Appendix 2).

SAMPLES REQUIRED FOR DETECTION

Qualitative sampling (as described in Clarkson 1996) over a 4-year period (1995–1998) enhanced detections of species (Table 10). Most of these supplementary detections occurred during years when only 1 or 2 stations were sampled and involved species that occur at low densities.

The following equation, based on the Poisson distribution, can be used to estimate the number of stations to be sampled (n) while risking a specified probability of failure to detect (β) for fish at various densities (m). Green and Young (1993) validate the equation for densities below 0.1 per sample unit.

$$n = -\frac{1}{m}\ln\beta$$

For instance, if researchers don't want to miss a particular species more than 5% of the time,

$$-\ln(\beta) \approx 3$$

so that

$$n = \frac{3}{m}$$

Note that this formula predicts a 95% likelihood of detecting fish …

in x stations	if minimum density of fish per station is
1	3.00
2	1.50
3	1.00
4	0.75

Comparison of these numbers with average densities in Table 9 indicates that for all species except CAIN, CYLU, and GAAF, detection of the species in at least one reach it currently occupies is problematic. This will also be true for any species in any reach where they newly occur at low densities.

CONCLUSIONS

ABILITY TO DETECT CHANGE

Although 9 of the 17 species currently in the monitored reaches can be effectively monitored to detect at least moderate annual trends, there are 5 species for which only extraordinary increases or decreases will be detectable. Only trends for CAIN and CYLU could be adequately described at each reach where they currently occur. Reasonable changes in the protocol would be required for description of these changes in the further 3 species. None of the 4 threatened or endangered species addressed in the BO have been detected using the current protocol.

The results for specific tests of concordance indicate that an impact that propagates spatially, for instance if the CAP confluence acts as a point source for perturbations that spread throughout the community, might be difficult to detect on the San Pedro, since there is low consistency between reaches, anyway. Such a change might be easier to detect on the Gila, since we expect similar species associations between reaches each year, but describing the nature of the shift might be more difficult, because the Gila shows so much baseline variability between years anyway. Depending on the needs of the survey, separate power analyses could be constructed for each river.

If there is an impact of the CAP canal, we might expect that reaches that are closer to the CAP confluence with the Gila River will show any impacts first. This could be simulated, for instance, by allowing for the full estimated population decline, but putting this decline on the 3 reaches that are closest to the CAP canal (similar to Beier and Cunningham 1996). This is not the only possible scenario under which some reaches might show impacts while others do not. Another approach might use the inverse of distance from the point source as a surrogate for area in species-area formulations (Nichols and others 1998). Before these sort of specific scenarios are investigated, the monitoring agencies would have to decide exactly which scenarios are of concern.

After the data are collected, for analyses to detect different impacts at different reaches, I recommend an ANCOVA that includes all reaches and a time-by-reach interaction term, rather than performing separate analyses for each reach. If an initial test for autocorrelation indicates it is significant (Durbin-Watson test; Neter and others 1985), a simple transformation of the counts can help adjust the data to be more appropriate for the usual ANCOVA model (Neter and others 1985). (After the analysis, the slope and its confidence interval can be back-transformed to see whether the interval includes zero, and the trend is considered not statistically significant.) Regardless of whether autocorrelation is significant, a hierarchical approach beginning with a model that examines how abundance is an outcome of a given year, a given reach, and the

interactions between years and reaches will answer the primary questions. First, if the interaction between year and reach is significant, this will indicate that relative frequencies are changing in at least some reaches, and not all reaches are behaving the same way. Graphical examination and statistical post-hoc tests can reveal which reaches are changing. If there is no significant reach-by-time interaction, but the effect of time is significant, then the system as a whole is reflecting the same trend.

RECOMMENDED CHANGES TO THE SAMPLING PROTOCOL

Conducting power analyses during the pilot phase of a study is critical because it allows a program to meet its stated goals. Indeed, Clarkson (1998) calls for the management action plan to "… define threshold criteria for fish species richness, distribution, and assemblage structure indices…." At this point, where goals are not well-specified, a power analysis also provides the opportunity to decide which goals would be realistic.

The basic power analysis leads to the recommendation that 3 stations per reach be retained, or that two 400-m stations be sampled per reach, if power considerations for ICPU, LECY, MISA, and PIPR are important. Also, because a higher Type I error level would considerably improve power analyses for some species (α=0.10 for CYCA, ICPU, and LECY; α=0.20 for MISA), this is the time to consider the relative consequences of Type I and II errors for detecting trends in each species.

Considerable changes to the protocol would be required in order to adequately describe any changes in abundance of other species. One option might be to use stratified sampling to reduce the CV for counts. The premise of stratified sampling is that if a species utilizes some habitat types (pools, riffles, or runs, for instance) in preference to others, the variability in counts will be lower if habitat use is taken into account. The feasibility of this approach, which would make use of data already collected under the monitoring protocol, would have be evaluated in a separate analysis of habitat use by each species.

In the absence of stratified sampling, in order to describe trends for some of the less common or more variable species in the monitored reaches, the sample sizes would have to be increased. Green (1979) presents a formula for sample sufficiency that can be used in a simplified form to begin answering this question. The limiting factor for the current project is sample sufficiency for uncommon species. Focussing on these species, we can move to use of the Poisson distribution for describing abundance.

The formula Green uses is:

$$CV = \sqrt{\frac{n}{T^2} s^2}$$

That is, precision at each reach (*CV*) is a function of the number of stations (*n*), the total number of individuals sampled in each reach (*T*), and the variance in *T* (s^2). The variance can also be expressed as a function of the mean density (*m*) using Taylor's power law (Taylor 1961):

$$s^2 = am^b$$

Substituting and rearranging, Green (1979) arrives at:

$$\log(T_n) = \left[\frac{\log\left(CV^2 / a\right)}{b-2} \right] + \frac{b-1}{b-2} \log n$$

Fortunately for the purposes of this report, counts that are distributed in a Poisson manner, so the variance equals the mean, *b*=*a*=1, and the above formula reduces to

$$\log(T_n) = -\log\left(CV^2\right)$$

Considering the variables that remain, for uncommon species, the precision of density estimates (*CV*) is not a function of the number of stations in each reach taken (nor is it directly a function of the length of the station), but of the total number of each species that is collected in each reach. Designing sampling protocols using species counts instead of linear sample size has also been described in Wilson (1996). This is the minimum number of fish of each species to collect in each reach to achieve a particular level of variability. Figure 13 illustrates how the number of fish to sample increases with increasing levels of required precision.

If it is unpalatable to decide the sampling intensity on the fly in the field, is to increase the total length sampled in each reach. Using this approach, and assuming that average counts will increase proportionately with increasing sample length, we can arrive at an estimate of the number of stations and/or reaches we would have to sample to achieve a specific precision (Table 11). Angermeier and Smogor (1995) and Hankin (1984) approached this same issue by considering the number of units of each type of habitat (pool, riffle, run) that should be sampled.

These calculations provide rough guidelines to use when balancing considerations of the relative risks of Type I and II errors, of the feasibility of decreasing variability in estimates, and in deciding on a minimum effect size (magnitude of trend) for the final analysis. These issues play off against one another because by increasing the length of the monitoring schedule and/or deciding on a larger than conventional Type I error, we could accommodate larger CVs (greater variability), and could accept smaller samples. Regardless of options that are considered, however, it seems clear (Table 11) that the status of species occurring at lowest densities will have to be evaluated by means other than by testing for trends in abundance.

Is it biologically meaningful if an uncommon species declines in relative density? As long as the species has not been locally extirpated, does any change from uncommon to even-more-uncommon really represent a change in status for the species? Similarly, if a common species increases 10% each year, this might be interpreted as a meaningful change in the community, but if an uncommon species increases by this amount, it will still be very uncommon. For uncommon species, instead of tracking changes in density, Strayer (1999) advocates tracking any changes in the number of sites where these species are present. However, the power of this technique is limited by the number of sites that are followed, and the current protocol indicates that the 7 reaches currently monitored represent the upper limit in the number of sites.

DESCRIBING SPECIES ASSEMBLAGES

Evaluating change in relative frequency

As argued by Rahel (1990), relative rankings may be less variable than absolute counts. Any decrease in variability makes it easier to detect an underlying trend. Should the options of analyses done on relative frequencies be considered? Yant and others (1984) argue against this approach because estimates of proportions will differ in accuracy depending on the sample size. Smith (1980) also mentions that ratio indices have bias inversely proportional to the number of samples. Even if the protocol moved to sampling approximately the same number of fish each time, use of relative frequencies implies a further assumption, that when absolute abundance changes, all species are affected proportionately. Analysis of percents is apparently done to control for sample size, on the assumption that in larger samples, each species will increase proportionately. However, if there are one-tenth as many fish in one sampling period compared to another, the decline may be expected to impact some species more than others, so the sense that standardization has occurred my be an illusion. If this assumption does not hold, for example, if species vary independently, use of percentages introduces statistical dependence on sample size for correcting for total sample size variation

Multispecies analyses

As described above, Green (1979) provides a multivariate approach to describing changes in species assemblages. Another alternative follows Legendre and others (1985). They describe an approach to treat shifts in species assemblages as steps, rather than smooth transitions. Their method clusters samples at the same site that are adjacent to one another in time. Large changes between samples are identified and treated statistically to identify discontinuities.

A further multivariate approach to visualizing shifts in multispecies assemblages is to develop believable data on whether species above a to-be-determined density are present. These data represent a description of a species assemblage, which can be examined using multidimensional analysis (MDA). Although a power analysis in the general sense can be applied to such an approach, for the MDA (and any other multivariate analysis), effect size will not seem familiar. Whereas for a single species, percent change is a meaningful effect measure, for an assemblage, definition of important effects will be more complex (Legendre and others 1985). The output of MDA cannot be expressed in familiar terms of "proportional change" or even in terms of "increases and declines." Instead, a separate metric will be necessary to describe patterns that might be of concern to management and watchdog agencies. This lack of direct, easily explicable connection to particular species is one drawback for management agencies attempting to simultaneously use multivariate descriptions and set conditions that trigger higher levels of attention or management actions. Whatever effect metric is chosen, in the case of MDA, power of the analysis comes from ability to ascertain presence or absence, leading discussion back to sample sufficiency for detection of rare species.

When rare species are a focal component of the assemblage

Hurlbert (1971) argued that indices of diversity or richness are inherently insensitive to rare species, because these indices are not developed to address non-numerical patterns such as functional (or conservation) importance. However, given limited sample sizes and resulting bias in such indices due to estimation of uncommon species, texts such as Ludwig and Reynolds (1988) caution against inclusion of uncommon species when computing correlation coefficients and proceeding with association analysis, among other things. Their general caution is that inclusion of uncommon species in analyses directed at species associations is expected to lead to spurious results.

Reanalyses of a study of assemblage persistence illustrates how conventional analyses of associations are hobbled by inclusion of rare species (Grossman and others 1982, Yant and others 1984, Rahel and others 1984). Kendall's W (used in the cited studies as well as some analyses in this report) does not weight consistency of common species more highly than of rare species, which are likely to be very inconsistent due to sampling artifacts. Therefore, in the original study (Grossman and others 1982) only the most common species were used for the analysis. The 10 most common species accounted for

at least 75% of the overall species numbers, and the authors concluded that the species assemblages had no inherent stability. However Yant and others (1984) used the same data set and reported that the assemblage was very stable when all species were considered. They concluded that stability of the association is entirely a function of the proportion of the association that is "uncommon," but there is no non-arbitrary way to determine which species to include in the analysis.

In essence, Yant and others (1984) return to Hurlbert's (1971) position that all species should be included in analysis of assemblages, but they justify this position from the opposing viewpoint. Because indices are sensitive to numerical abundance, any decision to eliminate uncommon species from the analysis will affect the index in an arbitrary fashion. This also summarizes the issues that face the current monitoring protocol. Currently, no distinction is made between species in terms of their importance value, so the task of describing changes in abundance of uncommon species remains a serious issue. If agencies involved in monitoring the Gila basin can agree that species that are currently relatively common are also of greatest interest for monitoring, single-species descriptions might be the most straightforward way to examine impacts of the CAP canal.

LITERATURE CITED

Angermeier PL, Schlosser IJ. 1989. Species-area relationships for stream fishes. Ecology 70:1450-62.

Angermeier PL, Smogor RA. 1995. Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. Canadian Journal of Fisheries and Aquatic Sciences 52:936-49.

Beier P, Cunningham SC. 1996. Power of track surveys to detect changes in cougar populations. Wildlife Society Bulletin 24(3):540-6.

Clarkson RW. 1996. Long-term monitoring plan for fish populations in selected waters of the Gila River basin, Arizona. 2 ed. Phoenix, Arizona: U.S. Bureau of Reclamation.

Clarkson RW. 1998. Results of fish monitoring of selected waters of the Gila River basin, 1995-1996. Phoenix, AZ: U.S. Bureau of Reclamation.

Cohen J. 1988. Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Debinski DM, Brussard PF. 1994. Using biodiversity data to assess species-habitat relationships in Glacier National Park, Montana. Ecological Applications 4(4):833-43.

Eberhardt LL, Simmons MA. 1992. Assessing rates of increase from trend data. Journal of Wildlife Management 56:603-10.

Gerrodette T. 1987. A power analysis for detecting trends. Ecology 68(5):1364-72.

Gibbs JP. 1995. Monitor: Users Manual. Department of Biology, Yale University, New Haven, Connecticut.

Gibbs JP, Droege S, Eagle P. 1998. Monitoring populations of plants and animals. Bioscience 48(11):935-40.

Green RH. 1989. Power analysis and practical strategies for environmental monitoring. Environ Res 50:195-205.

Green RH. 1979. Sampling design and statistical methods for environmental biologists. New York: Wiley.

Green RH, Young RC. 1993. Sampling to detect rare species. Ecological Applications 3(2):351-6.

Grossman GD, Moyle PB, Whitaker Jr. JO. 1982. Stochasticity in structural and functional characteristics of an Indiana stream fish assemblage: a test of community theory. American Naturalist 120:423-54.

Gryska AD, Hubert WA, Gerow KG. 1997. Use of power analysis in developing monitoring protocols for the endangered Kendall Warm Springs dace. North American Journal of Fisheries Management 17:1005-9.

Gurevitch J, Chester STJr. 1986. Analysis of repeated measures experiments. Ecology 67(1):251-5.

Hankin DG. 1984. Multistage sampling designs in fisheries research: applications in small streams. Canadian Journal of Fisheries and Aquatic Sciences 41:1575-91.

Harris RB. 1986. Reliability of trend lines obtained from variable counts. Journal of Wildlife Management 50:165-71.

Herbold B. 1984. Structure of an Indiana stream fish association: choosing an appropriate model. American Naturalist 124(4):561-72.

Hurlbert SH. 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology 52:577-86.

Kendall KC, Metzgar LH, Patterson DA, Steele BM. 1992. Power of sign surveys to monitor populations trends. Ecological Applications 2(4):422-30.

Kovalak WP, Dennis SD, Bates JM. 1986. Sampling effort required to find rare species of freshwater musselsIsom BG, editor. Rationale for Sampling and Interpretation of Ecological Data in the Assessment of Freshwater Ecosystems. Philadelphia: American Society for Testing and materials. p 34-45.

Legendre P, Dallot S, Legendre L. 1985. Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. American Naturalist 125:257-88.

Ludwig JA, Reynolds JF. 1988. Statistical Ecology. New York: John Wiley and Sons.

Lyons J. 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. North American Journal of Fisheries Management 12:198-203.

Manly BFJ. 1997. Randomization, bootstrap and Monte Carlo methods in biology. 2 ed. London: Chapman and Hall.

Millard SP, Lettenmaier DP. 1986. Optimal design of biological sampling programs using the analysis of variance. Estuarine, Coastal and Shelf Science 22:637-56.

Moyle PB, Vondracek B. 1985. Persistence and structure of the fish assemblage in a small California stream. Ecology 66:1-13.

Neter J, Wasserman W, Kutner MH. 1985. Applied linear statistical models. 2 ed. Homewood, Illinois: Irwin.

Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR. 1998. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology 12(6):1390-8.

Paller MH. 1995. Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. North American Journal of Fisheries Management 15:110-20.

Peet RK. 1975. Relative diversity indices. Ecology 56:496-8.

Peterman RM. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47:2-15.

Peterman RM, Bradford MJ. 1987. Statistical power of trends in fish abundance. Canadian Journal of Fisheries and Aquatic Sciences 44:1879-89.

Peterman RM, Routlege RD. 1983. Experimental management of Oregon coho salmon (*Oncorhynchus kisutch*): designing for yield of information. Canadian Journal of Fisheries and Aquatic Sciences 40:1212-23.

Rahel FJ. 1990. The hierarchical nature of community persistence: a problem of scale. American Naturalist 136(3):328-44.

Rahel FJ, Lyons JD, Cochran PA. 1984. Stochastic or deterministic regulation of assemblage structure? It may depend on how the assemblage is defined. American Naturalist 124:583-9.

Siegel, S., and NJ. Castellan, Jr. 1988. Nonparametric statistics for the behavioral sciences. 2nd edition. New York: McGraw Hill. 399 p.

Simonson TD, Lyons J. 1995. Comparison of catch per effort and removal procedures for sampling steam fish assemblages. North American Journal of Fisheries Management 15:419-27.

Smith SJ. 1980. Comparison of two methods of estimating the variance of the estimate of catch per unit effort. Canadian Journal of Fisheries and Aquatic Sciences 37:2346-51.

Strayer DL. 1999. Statistical power of presence-absence data to detect population declines. Conservation Biology 13(5):1034-8.

Taylor LR. 1961. Aggregation, variance and the mean. Nature 189:732-5.

US Fish and Wildlife Service. 1994. Endangered Species Act Section 7 Biological Opinion on transportation and delivery of Central Arizona Project water to the Gila River Basin (Hassayampa, Aqua Fria, Salt, Verde, San Pedro, Middle and Upper Gila River and associated tributaries) in Arizona and New Mexico. 2-21-90-F-119, April 15, 1994

van Strien AJ, van de Pavert R, Moss D, Yates TJ, van Swaay CAM, Vos P. 1997. The statistical power of two butterfly monitoring schemes to detect trends. J. Applied Ecology 34:817-28.

Wilson JR. 1996. Power analysis of fish monitoring data in Arizona. in: Clarkson RW. Long-term monitoring plan for fish populations in selected waters of the Gila River basin, Arizona. 2 ed. Phoenix, Arizona: U.S. Bureau of Reclamation.

Yant PR, Karr JR, Angermeier PL. 1984. Stochasticity in stream fish communities: an alternative interpretation. American Naturalist 124:573-82.

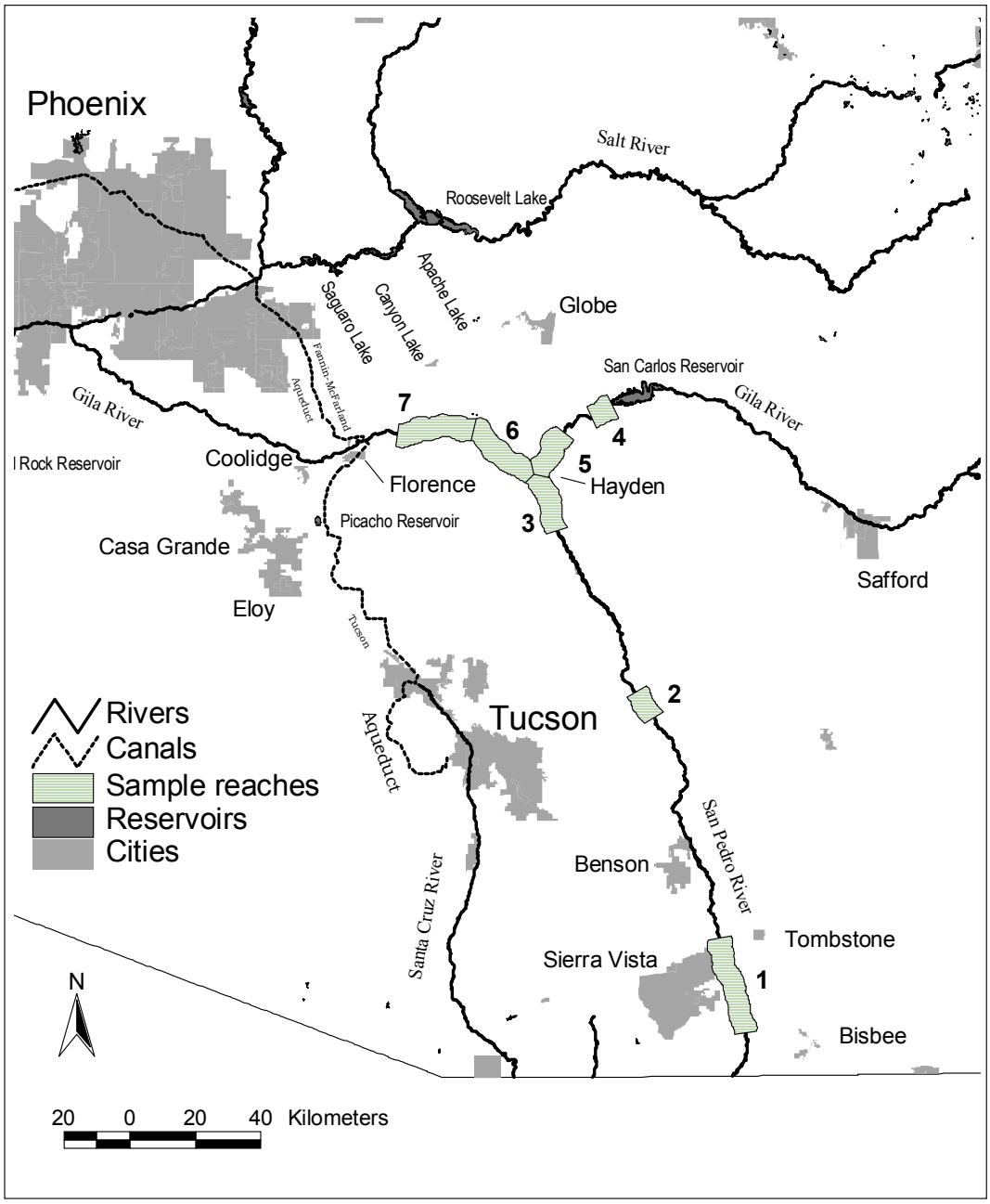

Figure 1. Map of sampled reaches on the Gila and San Pedro Rivers. Reaches are identified in the figure and in the text using the same numbers. CAP canals are shown; Salt River Project canals in the Phoenix metro area (not shown) join the CAP canals near the crossing of the Salt River.

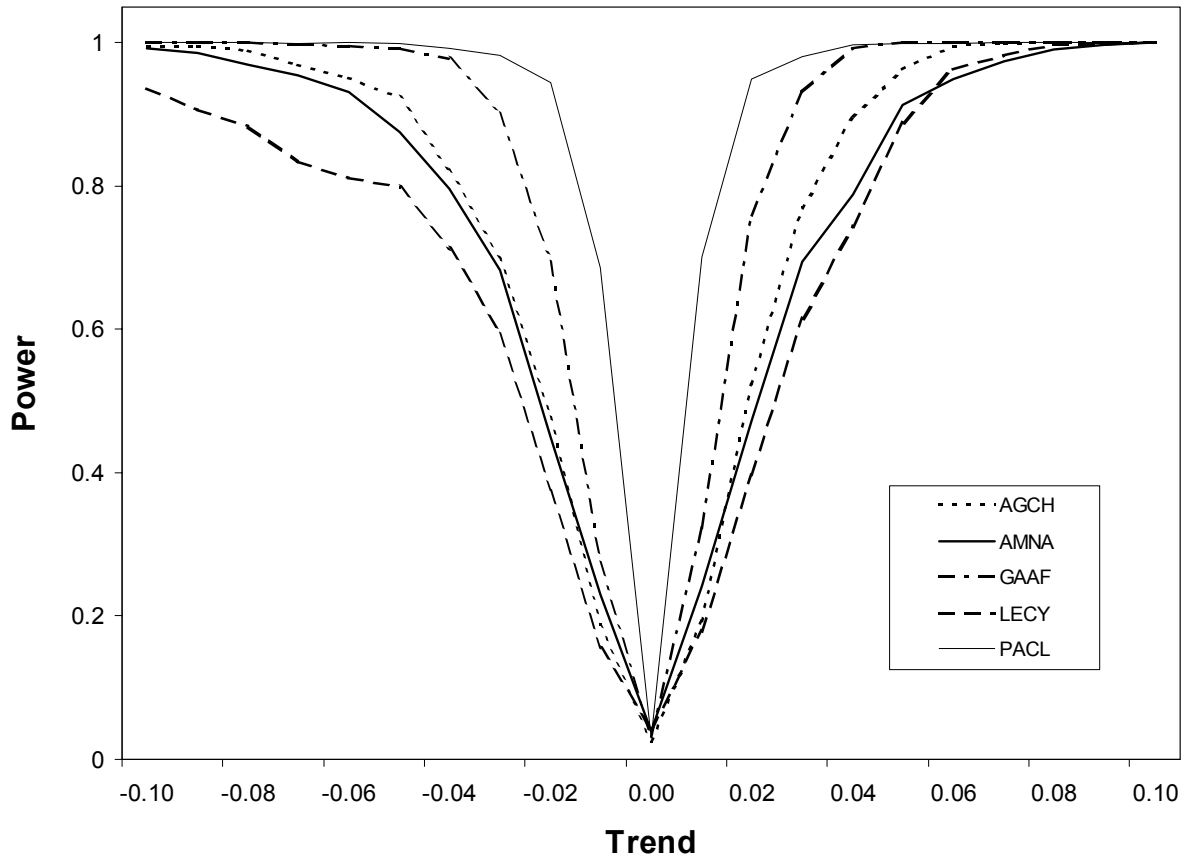

Figure 2. Power to detect trends in abundance for species currently found in all seven monitored reaches.

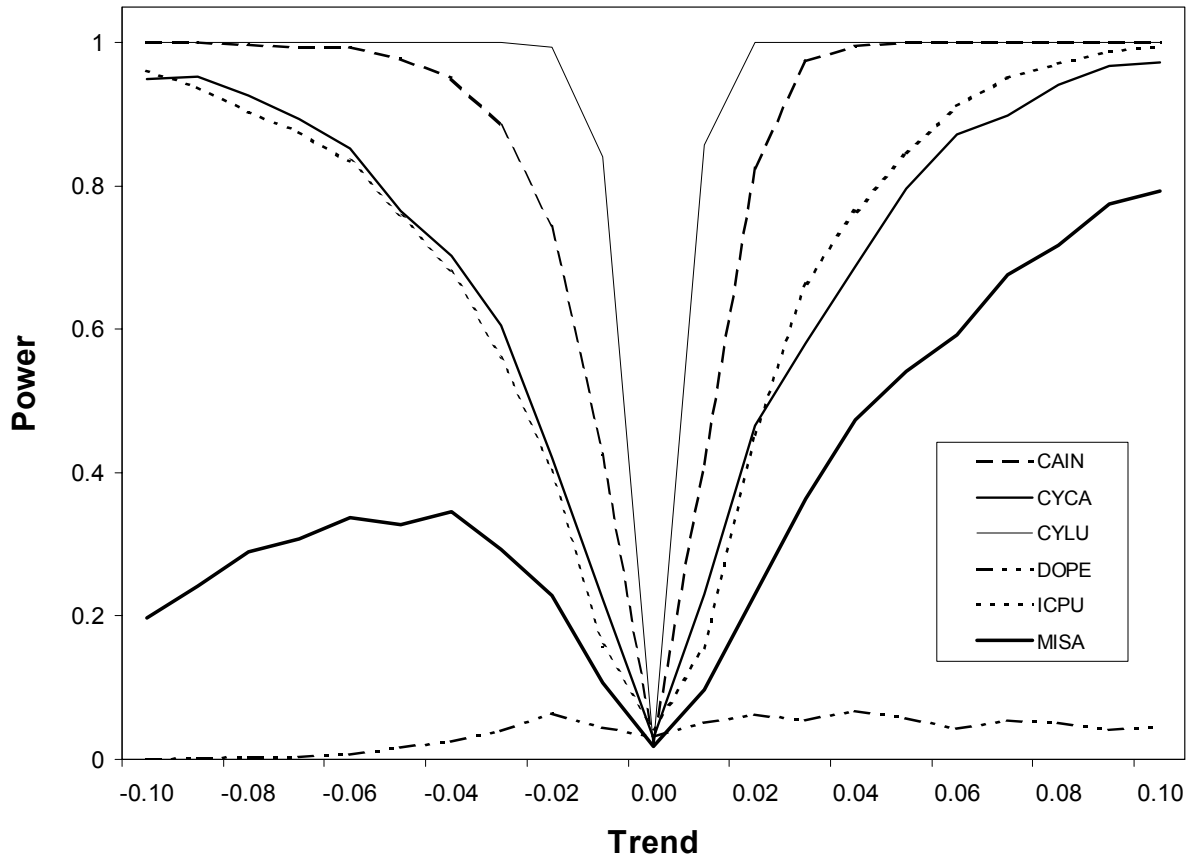

Figure 3. Power to detect trends in abundance for species currently found throughout the Gila River and on the San Pedro at the confluence with the Gila River (reaches 3 through 7; MISA is found in reaches 1 and 4 through 6).

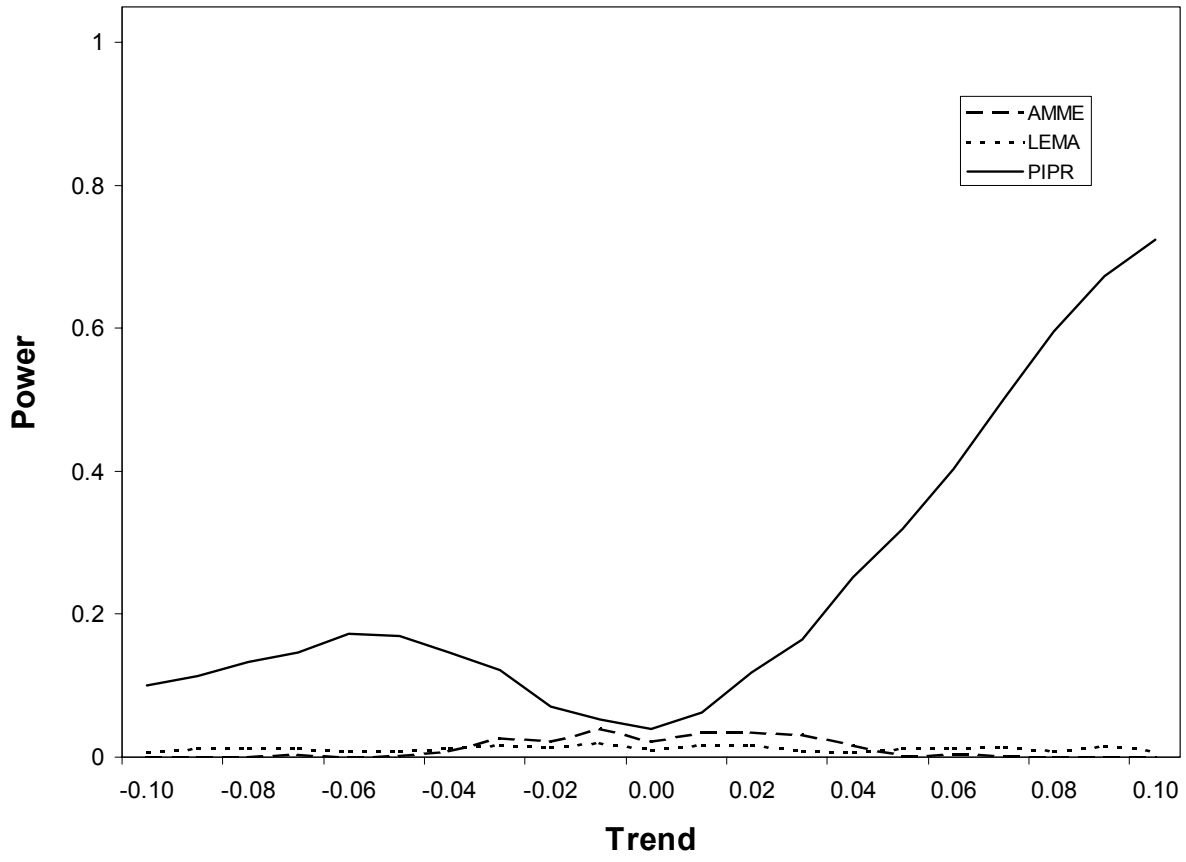

Figure 4. Power to detect trends for species currently found only on the San Pedro River.

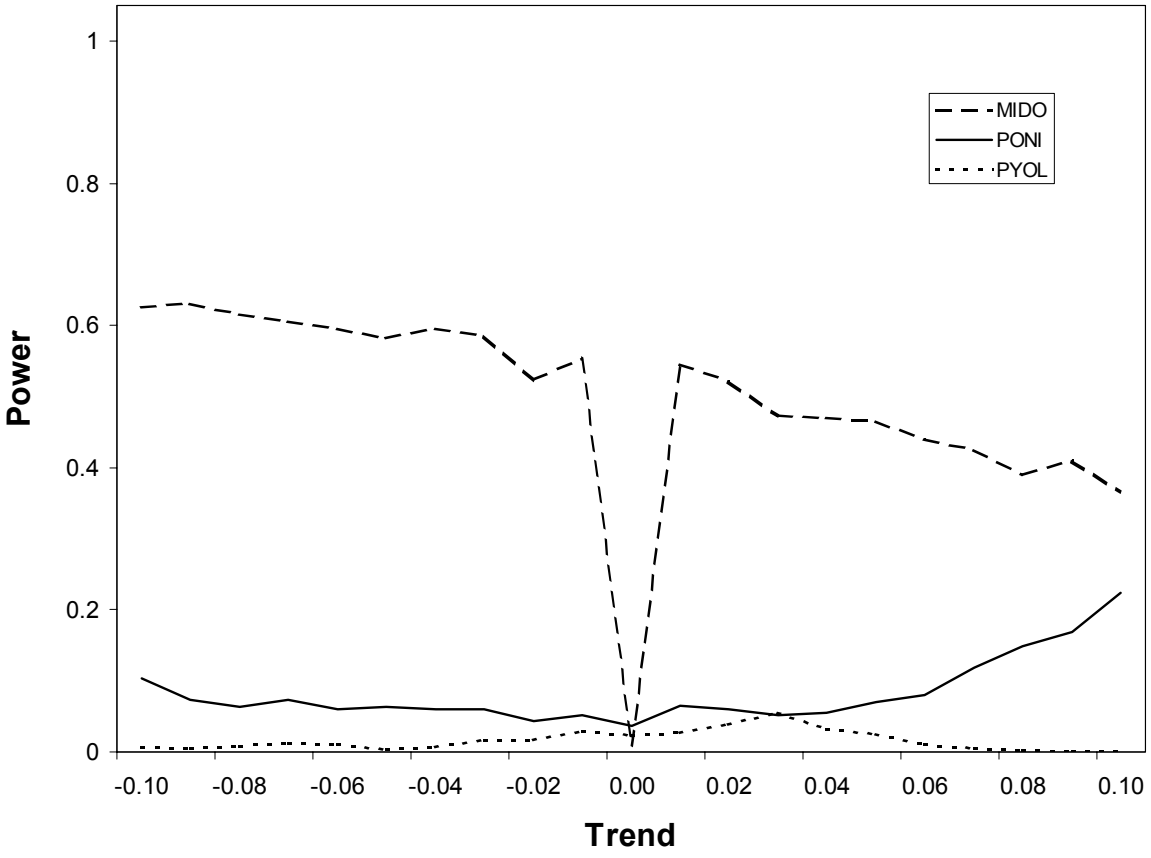

Figure 5. Power to detect trends in abundance for species found only on the Gila River in reaches 4 and 5, closest to the San Carlos Reservoir.

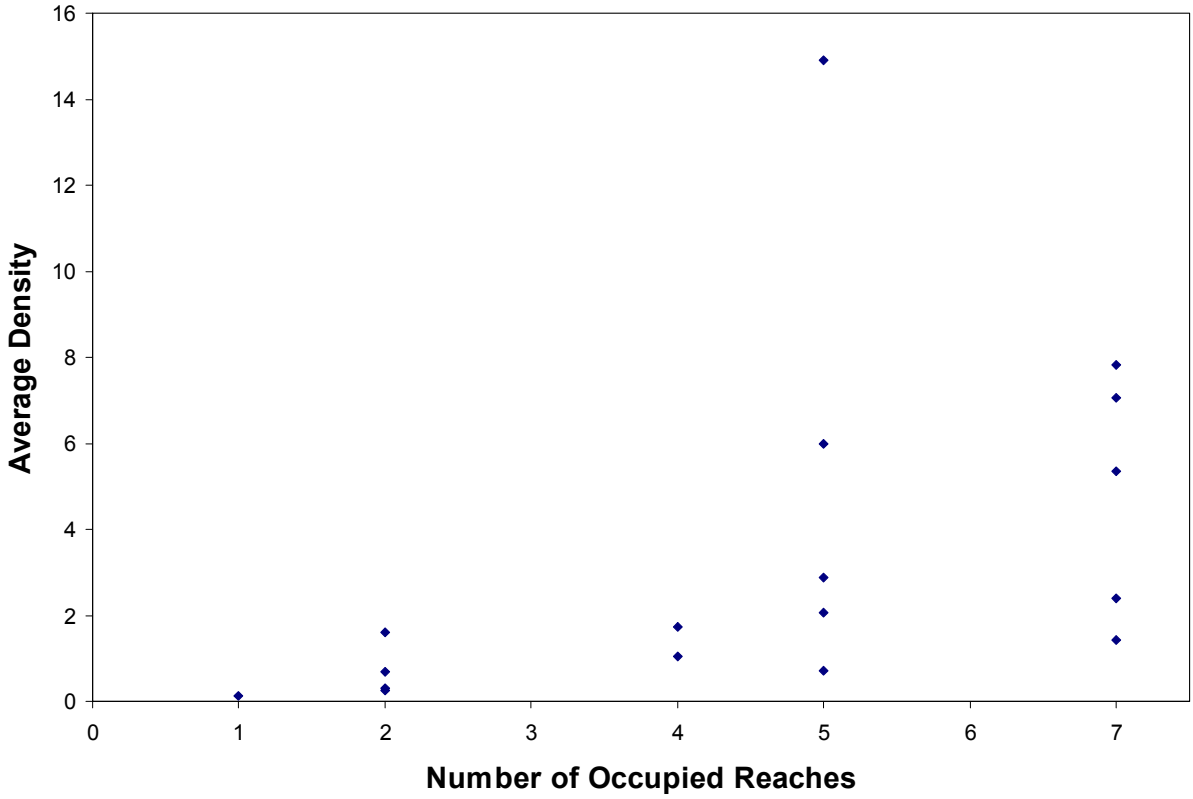

Figure 6. Relationship between spatial distribution and average density.

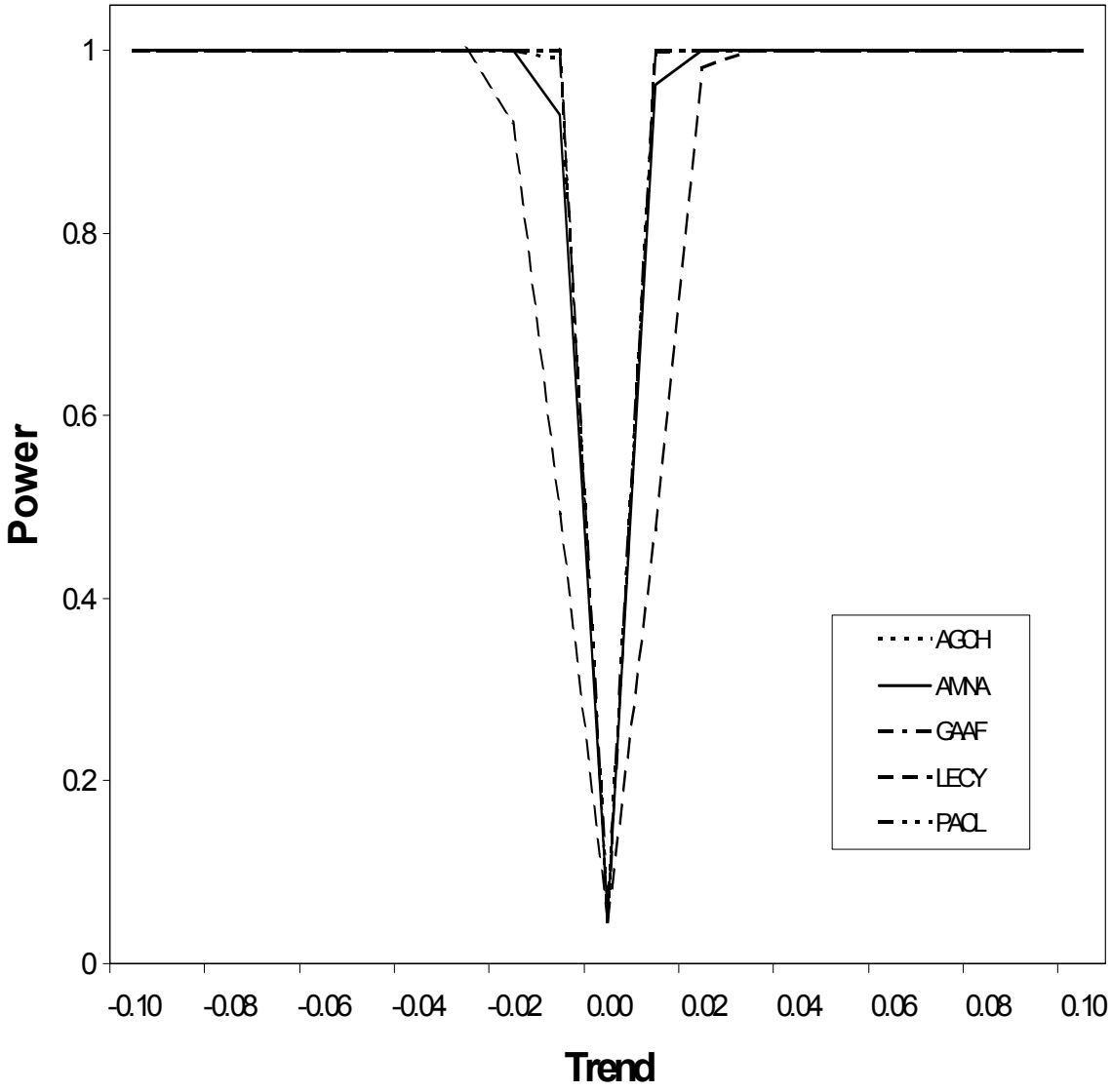

Figure 7. Power analysis for ability to detect a trend in abundance in the single reach with the highest density of each species. These species are currently found in all seven reaches (compare to Figures 2 and 8).

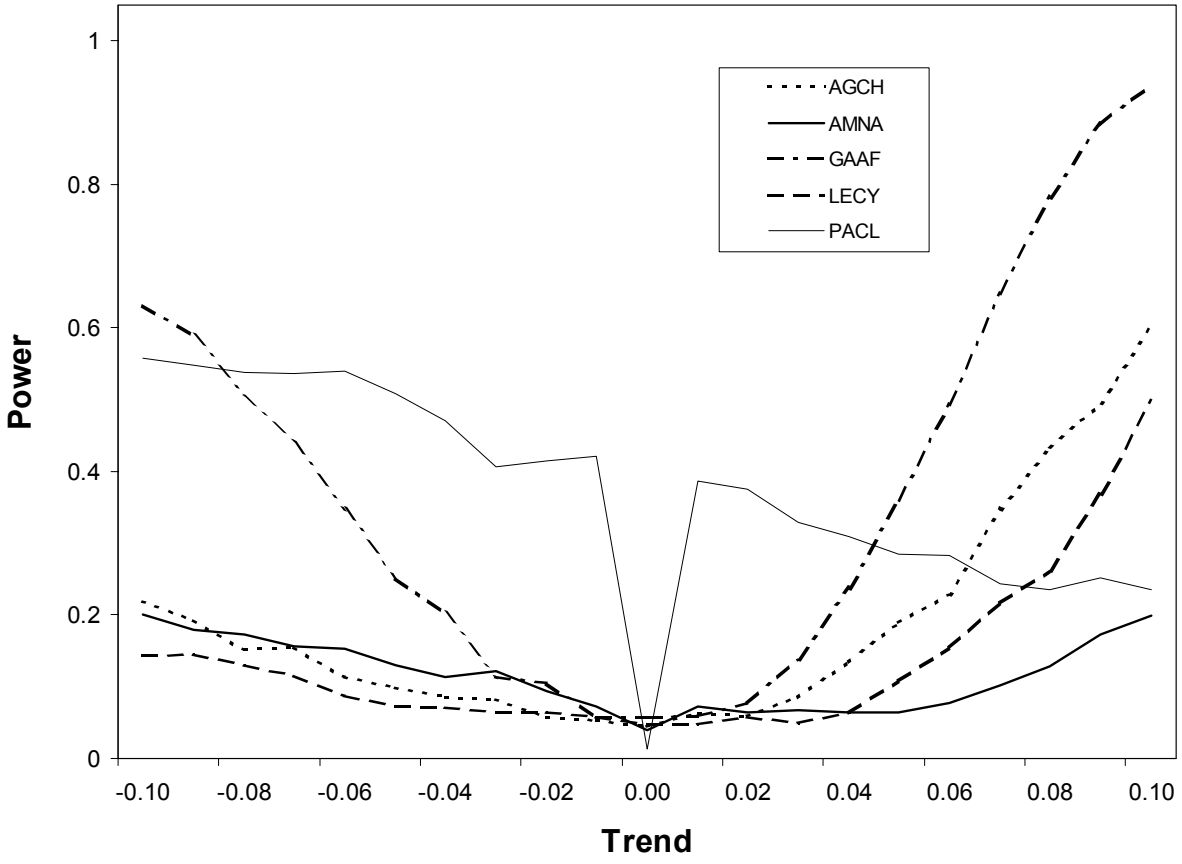

Figure 8. Power analysis for ability to detect a trend in abundance in the single reach with the lowest density of each species. These species are currently found in all seven reaches (compare to Figures 2 and 7).

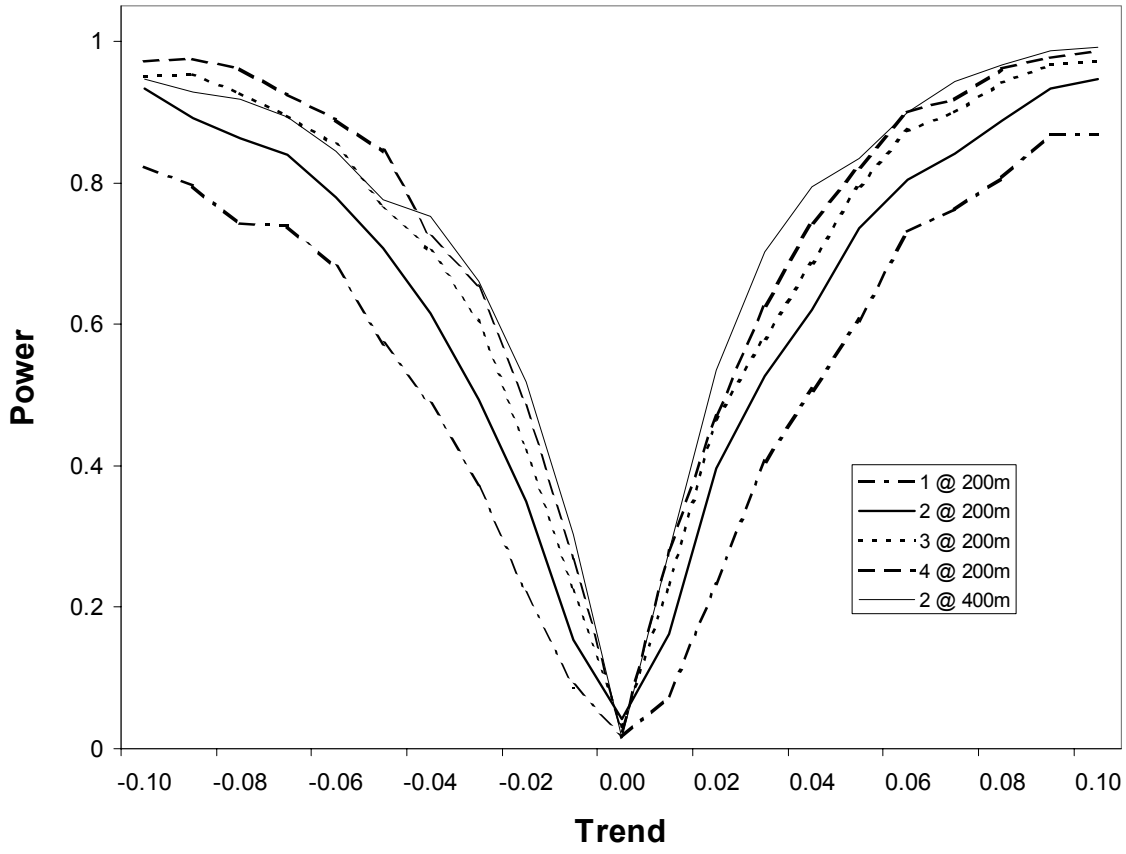

Figure 9. Effect of effort level on power to detect trends in CYCA. Effort is described by number of stations at a reach and by length of river sampled at each station.

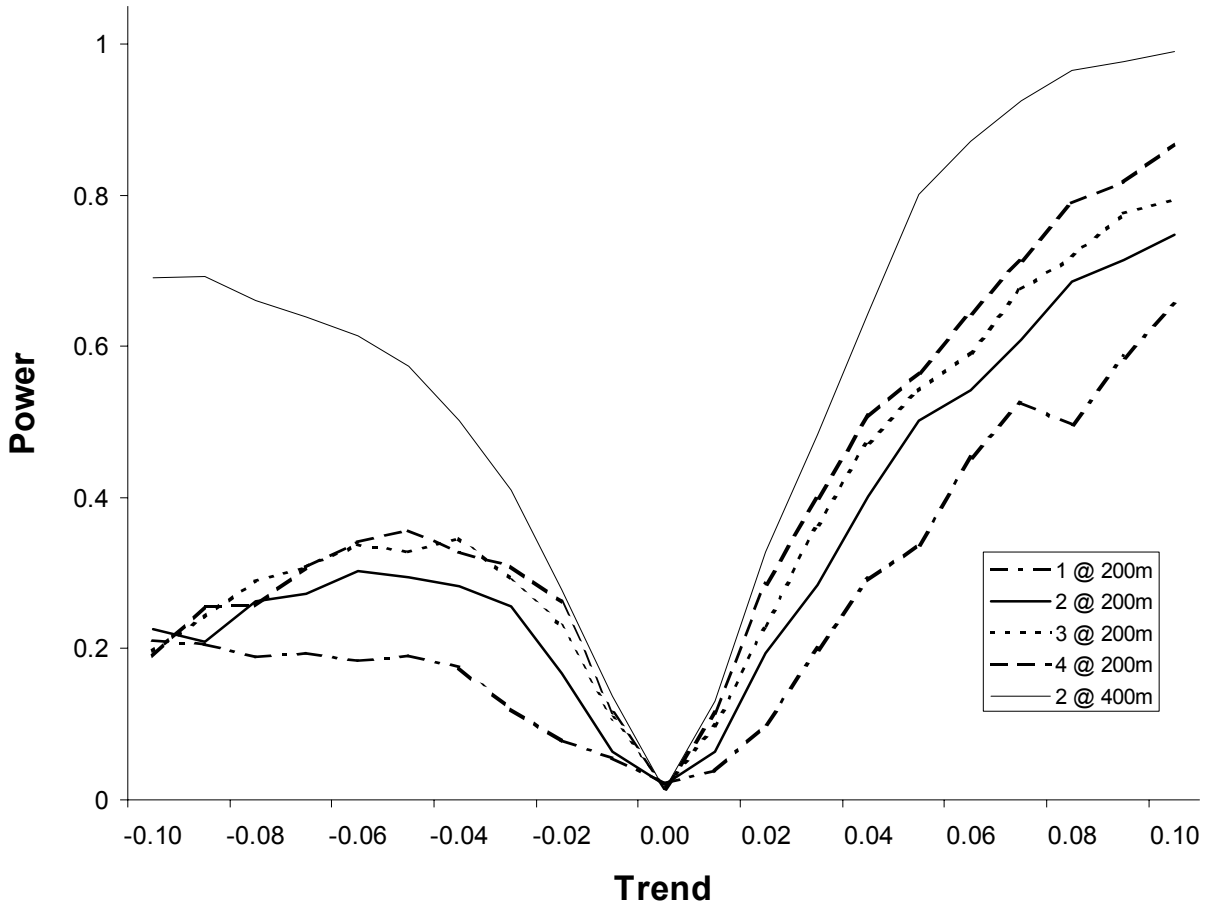

Figure 10. Effect of effort level on power to detect trends in MISA. Effort is described by number of stations at a reach and by length of river sampled at each station.

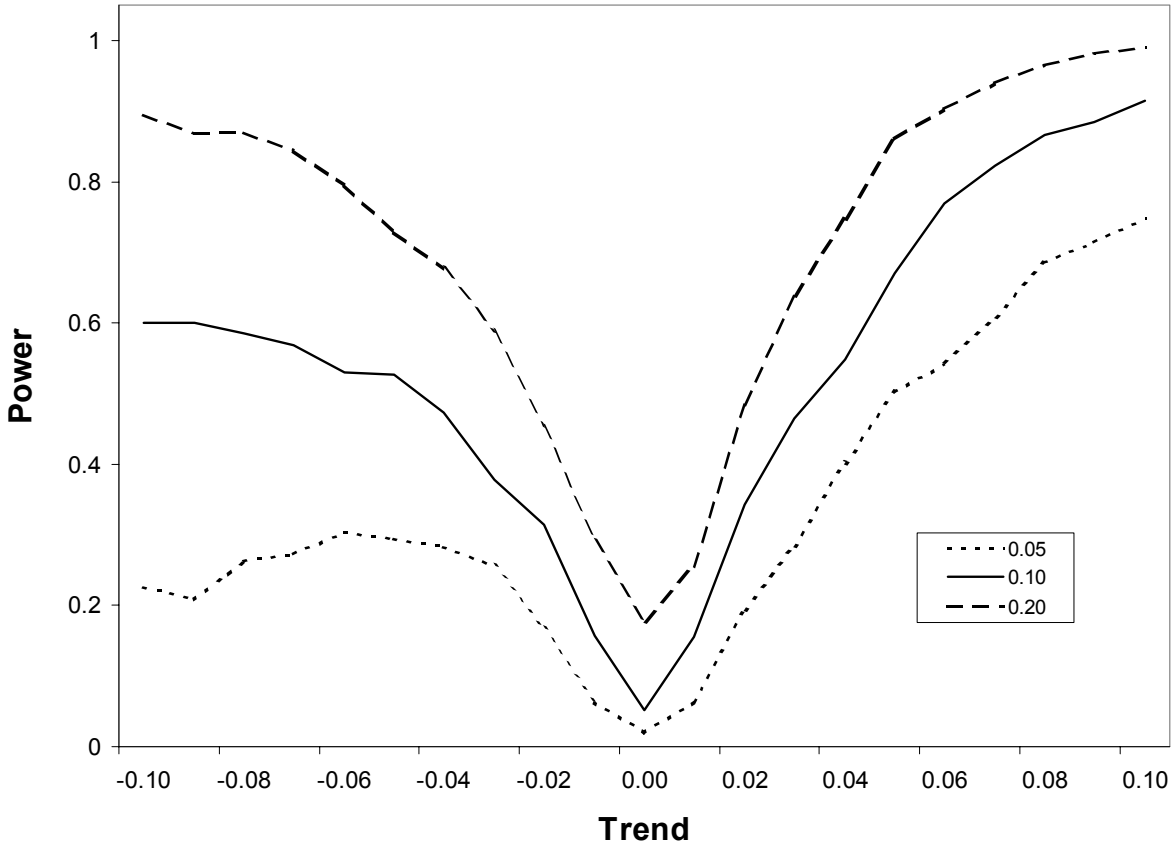

Figure 11. Effect of accepting different Type I error probabilities on power to detect trends in MISA.

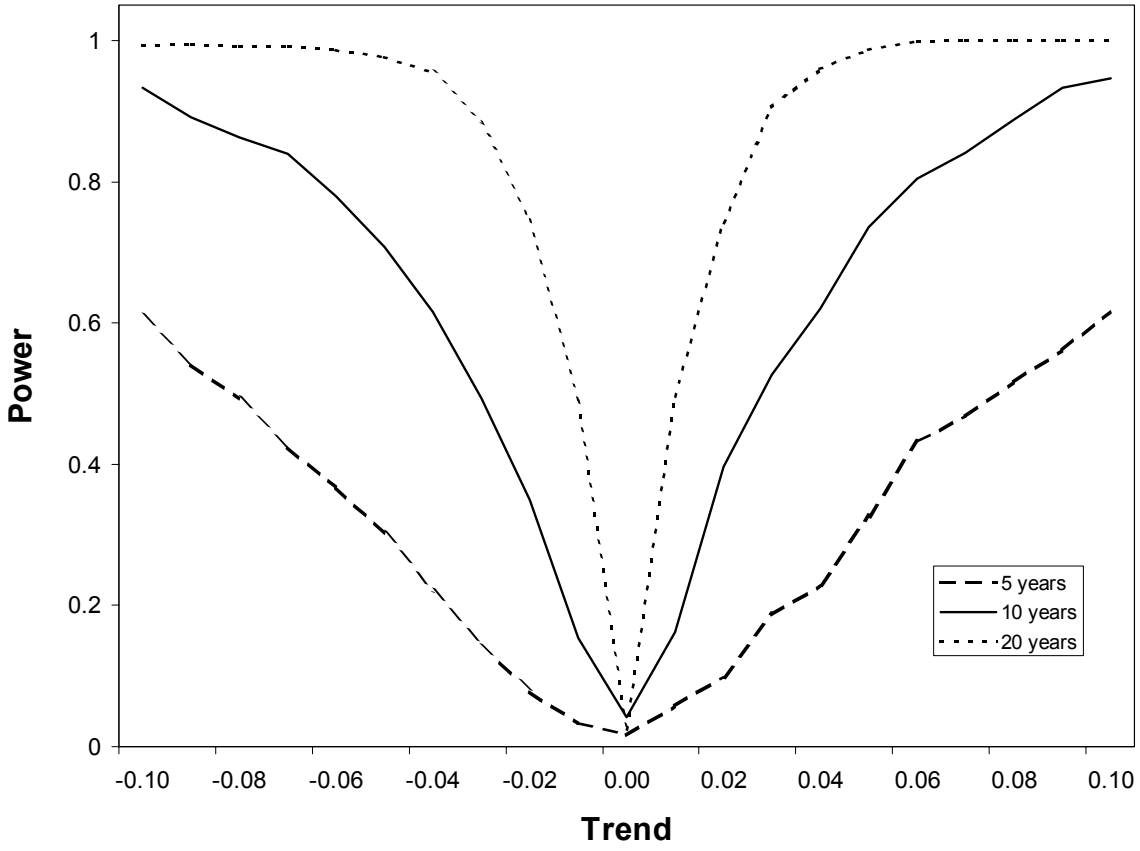

Figure 12. Effect of monitoring for different time periods on power to detect the same annual exponential trend in CYCA.

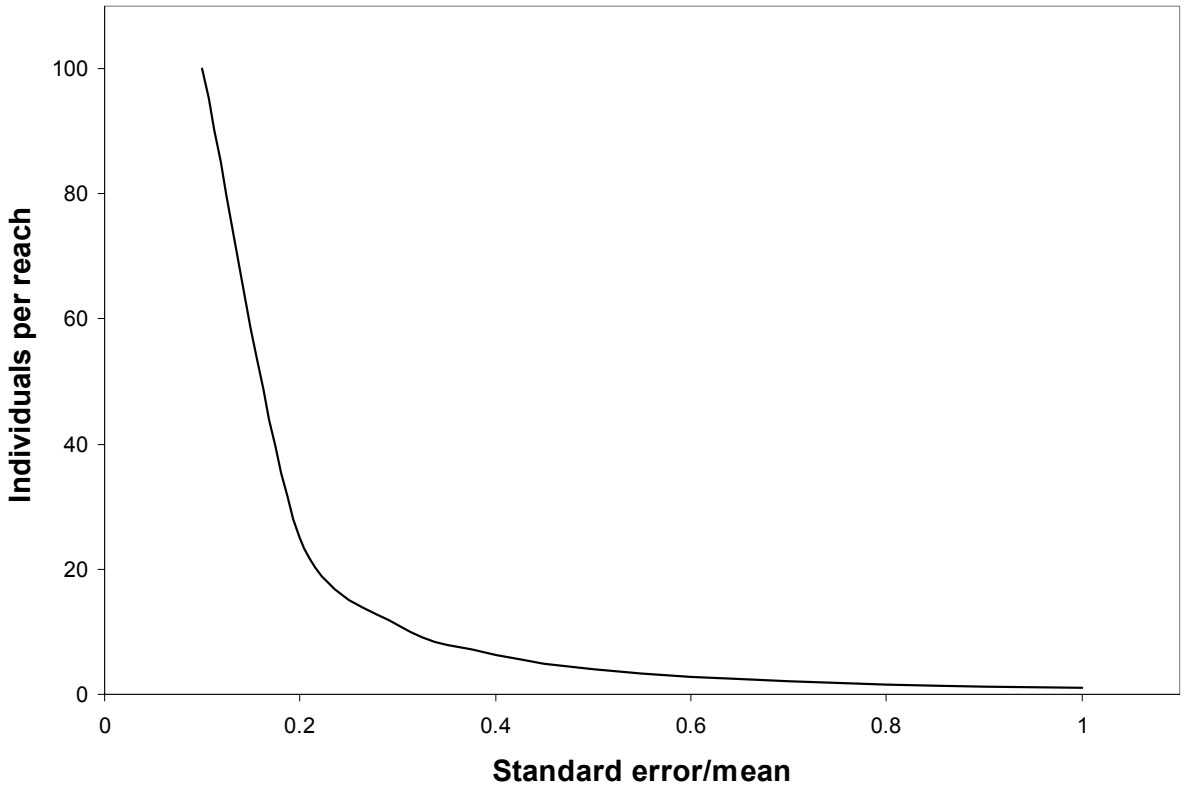

Figure 13. Number of individual fish of each species to sample per reach in order to achieve a given standard deviation of the density estimate.

Table 1. Latin and common names of fish and acronyms used in the document.

Acronym	Latin name	Common name
AGCH	*Agosia chrysogaster*	Longfin dace
AMME	*Ameiurus melas*	Black bullhead
AMNA	*Ameiurus natalis*	Yellow bullhead
CAIN	*Catostomus insignis*	Sonora sucker
CYCA	*Cyprinus carpio*	Common carp
CYLU	*Cyprinella lutrensis*	Red shiner
DOPE	*Dorosoma petenense*	Threadfin shad
GAAF	*Gambusia affinis*	Mosquitofish
ICPU	*Ictalurus punctatus*	Channel catfish
LECY	*Lepomis cyanellus*	Green sunfish
LEMA	*Lepomis macrochirus*	Bluegill
MEFU	*Meda fulgida*	Spikedace
MIDO	*Micropterus dolomieu*	Smallmouth bass
MISA	*Micropterus salmoides*	Largemouth bass
PACL	*Pantosteus clarki*	Desert sucker
PIPR	*Pimephales promelas*	Fathead minnow
PONI	*Pomoxis nigromaculatus*	Black crappie
PYOL	*Pylodictis olivaris*	Flathead catfish

Table 2. Proportional change from initial population size after several years of experiencing a particular annual trend.

Annual trend	Years monitored		
	5	10	20
+0.10	1.46	2.36	6.12
+0.09	1.41	2.17	5.14
+0.08	1.36	2.00	4.32
+0.07	1.31	1.84	3.62
+0.06	1.26	1.69	3.03
+0.05	1.22	1.55	2.53
+0.04	1.17	1.42	2.11
+0.03	1.13	1.30	1.75
+0.02	1.08	1.20	1.46
+0.01	1.04	1.09	1.21
0	1.00	1.00	1.00
–0.01	0.96	0.91	0.83
–0.02	0.92	0.83	0.68
–0.03	0.89	0.76	0.56
–0.04	0.85	0.69	0.46
–0.05	0.81	0.63	0.38
–0.06	0.78	0.57	0.31
–0.07	0.75	0.52	0.25
–0.08	0.72	0.47	0.21
–0.09	0.69	0.43	0.17
–0.10	0.66	0.39	0.14

Table 3. Scenarios explored for each species using Program Monitor simulations. Trends were explored that reflected proportional change of –0.1 to +0.1 in increments of 0.01.

Scenario	Stations per reach	Years sampled	Station length (meters)	Probability of Type I error	Reaches simulated
1	3	10	200	0.05	All
2	1	10	200	0.05	All
3	2	10	200	0.05	All
4	4	10	200	0.05	All
5	2	10	100	0.05	All
6	2	10	400	0.05	All
7	2	10	200	0.10	All
8	2	10	200	0.20	All
9	2	5	200	0.05	All
10	2	20	200	0.05	All
13	3	10	200	0.05	Most dense
14	3	10	200	0.05	Least dense

Table 4. History of sampling at each station. 0=Not Sampled. 1=Non-quantitative sampling. 2=Quantitative sampling following Clarkson (1996). Reach numbers as in Figure 1. Station names taken from Clarkson (1996).

River	Reach	Station Name	Year							
			1991	1992	1993	1994	1995	1996	1997	1998
San Pedro	0	SAN MANUEL	1	1	1	1	0	0	0	0
San Pedro	1	HEREFORD	0	0	0	0	2	2	2	2
San Pedro	1	LEWIS SPR	0	0	0	1	2	2	2	2
San Pedro	1	CHARLESTON	0	0	0	1	2	2	2	2
San Pedro	2	HUGHES RANCH	1	1	1	1	2	2	2	2
San Pedro	2	SOZA RANCH	0	0	0	0	2	2	2	2
San Pedro	3	ARAVAIPA	1	1	1	1	2	2	2	2
San Pedro	3	DUDLEYVILLE	1	1	1	1	2	2	2	2
San Pedro	3	MOUTH	0	0	0	0	2	2	2	2
Gila	4	COOLIDGE	1	1	1	1	2	0	2	2
Gila	4	HOOK and LINE	0	0	0	0	1	0	2	2
Gila	5	DRIPPING SPR	0	0	0	0	2	0	2	2
Gila	5	CHRISTMAS	1	1	1	1	0	0	2	2
Gila	5	OCARROLL	0	0	0	0	2	0	2	2
Gila	6	SAN PEDRO	1	1	1	1	0	0	2	2
Gila	6	KEARNEY	0	0	0	0	0	0	2	2
Gila	6	RIVERSIDE	1	1	1	1	0	0	2	2
Gila	7	DIAMOND A	1	1	1	1	2	0	1	0
Gila	7	COCHRAN	1	1	1	1	1	2	2	2
Gila	7	BOX O WASH	1	1	1	1	2	0	2	2

Table 5. Species present in each reach or each year in each river.

	AGCH	AMME	AMNA	CAIN	CYCA	CYLU	DOPE	GAAF	ICPU	LECY	LEMA	MIDO	MISA	PACL	PIPR	PONI	PYOL
Species present at least one year in all reaches																	
Gila R.	x	x		x	x	x	x	x	x	x				x			
San Pedro R.	x	x						x		x				x	x		
Species present in at least one reach each year																	
Gila R.	x	x		x	x	x			x					x			
San Pedro R.	x	x	x	x				x		x				x	x		

Table 6. Consistency of assemblages within reaches between years. Note that for Kendall's W, a significant p-value indicates concordance.

River	Reach	Consistency between years					
		Counts			Abundance rankings		
		χ^2	df	p-value	Kendall's W	df	p-value
San Pedro	1	95.3	18	0.000	0.87	7	0.001
	2	274.8	18	0.000	0.57	7	0.024
	3	1098.0	18	0.000	0.87	7	0.001
Gila	4	83.9	12	0.000	0.41	6	0.294
	5	324.6	12	0.000	0.61	6	0.088
	6	22.3	6	0.000	0.71	6	0.199
	7	416.8	18	0.000	0.52	6	0.053

Table 7. Consistency of assemblages across reaches. Only one station was sampled on the Gila River in 1996. Note that for Kendall's W, a significant p-value indicates concordance.

River	Year	Consistency between reaches					
		Counts			Abundance rankings		
		χ^2	df	p-value	Kendall's W	df	p-value
San Pedro	1995	1532.1	10	0.000	0.89	5	0.021
	1996	572.4	10	0.000	0.60	5	0.109
	1997	61.8	10	0.000	0.54	5	0.149
	1999	260.6	10	0.000	0.61	5	0.106
Gila	1995	164.3	12	0.000	0.71	9	0.024
	1997	184.0	24	0.000	0.52	9	0.027
	1998	452.8	24	0.000	0.51	9	0.033

Table 8. Serial correlation of counts for each species between all year-to-year pairs.

Species	Number of pairs	r	p-value
AGCH	15	0.77	0.001
AMME	6	0.60	0.203
AMNA	15	0.81	0.000
CAIN	9	0.83	0.005
CYCA	9	0.74	0.023
CYLU	9	–0.09	0.824
DOPE	6	1.00	0.000
GAAF	15	0.13	0.641
ICPU	9	0.51	0.164
LECY	15	0.90	0.000
LEMA	4	–0.36	0.639
MIDO	0	-	-
MISA	6	0.89	0.016
PACL	15	–0.10	0.723
PIPR	11	0.84	0.001
PONI	2	-	-
PYOL	2	-	-

Table 9. Coefficients of variation and mean counts at each reach. If a species was never detected in a reach, it was assumed not present and no counts were estimated. If a species was ever detected in any station of a particular reach, then for purposes of averaging, counts were set to zero at each station in that reach when the species was not detected.

Species	CV	San Pedro River reaches			Gila River reaches			
		1	2	3	4	5	6	7
AGCH	0.965	12.6	78.9	71.4	0.4	2.6	1.0	1.0
AMME	0.439	4.7	0.1					
AMNA	0.490	0.2	0.3	0.9	1.8	3.3	13.4	7.3
CAIN	0.546			1.6	1.9	30.2	5.7	6.5
CYCA	0.534			0.2	9.5	5.2	3.2	1.2
CYLU	0.716			5.7	9.6	36.4	6.1	30.2
DOPE	0.550				9.5	0.2	0.2	0.2
GAAF	0.754	36.5	5.8	9.8	1.4	1.3	3.5	2.6
ICPU	0.523			0.3	2.1	1.7	4.0	3.3
LECY	0.345	4.8	0.2	0.3	3.2	1.8	1.8	0.5
LEMA	0.290	0.1		0.6				
MIDO	0.135				0.1			
MISA	0.387	0.5			4.8	2.9	0.3	
PACL	0.705	20.1	0.1	5.0	3.5	12.6	6.3	15.6
PIPR	0.427	2.9	0.6	0.2	0.3		0.2	
PONI	0.204				0.3	0.3		
PYOL	0.223				1.6	0.1		

Table 10. Detections of species by efforts beyond quantitative sampling (Clarkson 1996).

Species	River	Reach	Year	Mean density in this reach	Stations sampled quantitatively this year
AMME	San Pedro	2	1996	0.1	1
AMME	San Pedro	2	1998	0.1	2
CAIN	Gila	4	1995	1.9	1*
DOPE	Gila	6	1997	0.2	3
GAAF	Gila	4	1995	1.4	1*
ICPU	Gila	7	1995	3.3	2*
LECY	San Pedro	2	1995	0.2	2
PACL	San Pedro	2	1995	0.1	2
PYOL	Gila	6	1998	Otherwise undetected	3

*An additional station was sampled qualitatively.

Table 11. Stations (200 m long) required per reach to achieve various coefficients of variation.

Precision	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
SPECIES	Station lengths									
AGCH	13	4	2	1	1	1	1	1	1	1
AMME	63	16	7	4	3	2	2	1	1	1
AMNA	42	11	5	3	2	2	1	1	1	1
CAIN	17	5	2	2	1	1	1	1	1	1
CYCA	35	9	4	3	2	1	1	1	1	1
CYLU	7	2	1	1	1	1	1	1	1	1
DOPE	96	24	11	6	4	3	2	2	2	1
GAAF	19	5	3	2	1	1	1	1	1	1
ICPU	49	13	6	4	2	2	1	1	1	1
LECY	71	18	8	5	3	2	2	2	1	1
LEMA	330	83	37	21	14	10	7	6	5	4
MIDO	817	205	91	52	33	23	17	13	11	9
MISA	59	15	7	4	3	2	2	1	1	1
PACL	15	4	2	1	1	1	1	1	1	1
PIPR	141	36	16	9	6	4	3	3	2	2
PONI	385	97	43	25	16	11	8	7	5	4
PYOL	147	37	17	10	6	5	3	3	2	2

Appendix 1. Fish counts at each station each year. A zero was entered if the species was undetected in a particular sample, but was present during a different year in that reach.

San Pedro River reaches

Reach	Station	Year	AGCH	AMME	AMNA	CAIN	CYCA	CYLU	DOPE	GAAF	ICPU	LECY	LEMA	MIDO	MISA	PACL	PIPR	PONI	PYOL
1	Hereford	1995	186	15	0					112		9	0			22	0		
1	Hereford	1996	21	19	0					149		9	0			19	0		
1	Hereford	1997	27	24	2					22		2	0			28	1		
1	Hereford	1998	20	27	0					1		9	0			50	0		
1	Lewis Spring	1995	0	3	0					65		20	0		2	25	2		
1	Lewis Spring	1996	0	17	0					31		16	0			12	6		
1	Lewis Spring	1997	6	2	1					85		11	0		6	7	0		
1	Lewis Spring	1998	4	8	0					87		1	0			0	0		
1	Charleston	1995	30	0	0					35		4	0		2	68	92		
1	Charleston	1996	8	7	0					76		0	0			133	33		
1	Charleston	1997	37	0	1					26		10	1		1	50	38		
1	Charleston	1998	167	0	0					62		5	0		1	63	7		
2	Hughes Ranch	1995	1390	1	0					422		0				0	0		
2	Hughes Ranch	1996	148	0	0					22		0				0	0		
2	Hughes Ranch	1997	30	0	6					4		0				0	2		
2	Hughes Ranch	1998	247	0	0					0		0				1	2		
2	Soza Ranch	1995	344	0	0					187		0				0	0		
2	Soza Ranch	1996	0	0	0					0		0				0	0		
2	Soza Ranch	1997	31	0	0					1		1				0	0		
2	Soza Ranch	1998	377	0	0					0		1				0	4		
3	Aravaipa	1995	5505		8	4	0	0		326	0	0				88	1		
3	Aravaipa	1996	524		2	4	0	6		6	0	0				82	0		
3	Aravaipa	1997	1		0	0	0	1		0	0	0				0	1		
3	Aravaipa	1998	24		0	0	0	0		1	0	0				0	0		
3	Dudleyville	1995	513		2	0	0	0		27	0	0				0	0		
3	Dudleyville	1996	179		0	3	1	32		14	1	1				27	0		
3	Dudleyville	1997	2		0	1	0	6		15	0	1				0	2		
3	Dudleyville	1998	15		0	1	3	28		1	2	2				6	0		
3	Mouth	1995	2184		3	0	0	0		324	0	0				0	0		
3	Mouth	1996	1136		1	26	0	61		10	0	0				346	0		
3	Mouth	1997	7		1	5	0	45		8	0	1				8	0		
3	Mouth	1998	12		5	2	0	30		9	4	0				0	0		

Appendix 1 (continued). Fish counts at each station each year. A zero was entered if the species was undetected in a particular sample, but was present during a different year in that reach.

Gila River reaches

Reach	Station	Year	AGCH	AMME	AMNA	CAIN	CYCA	CYLU	DOPE	GAAF	ICPU	LECY	LEMA	MIDO	MISA	PACL	PIPR	PONI	PYOL
4	Coolidge	1995	4		0	0	13	3	0	0	1	0		0	5	6	2	0	0
4	Coolidge	1997	0		3	0	25	2	1359	1	0	15		1	9	1	0	0	2
4	Coolidge	1998	0		0	0	18	17	65	0	15	1		0	3	0	0	0	8
4	Hook and Line	1997	0		3	16	47	94	4	13	5	9		0	9	22	0	1	2
4	Hook and Line	1998	0		27	31	0	43	2	6	3	16		0	7	10	0	1	3
5	Dripping Spring	1995	79		2	26	0	72	0	0	1	0			0	31		0	0
5	Dripping Spring	1997	0		2	64	78	0	0	96	3	3			14	3		0	0
5	Dripping Spring	1998	0		6	121	32	85	0	17	3	4			8	16		1	0
5	Christmas	1997	1		8	125	0	74	3	0	0	4			19	45		3	0
5	Christmas	1998	1		2	43	15	560	0	0	14	1			6	3		0	0
5	O'Carroll	1995	11		0	3	0	46	0	0	0	0			0	4		0	0
5	O'Carroll	1997	5		15	82	13	38	0	0	2	6			3	71		0	0
5	O'Carroll	1998	0		17	36	24	37	0	0	4	6			2	49		0	1
6	San Pedro	1997	4		10	29	10	51	0	2	2	12			0	30	2		
6	San Pedro	1998	10		26	23	16	103	2	50	4	0	16		1	105	0		
6	Kearney	1997	0		20	20	0	1	0	1	1	3			1	14	0		
6	Kearney	1998	0		37	1	14	1	0	13	35	2			0	2	0		
6	Riverside	1997	0		1	2	0	2	0	0	0	0			0	0	0		
6	Riverside	1998	0		18	0	1	1	0	1	14	2			0	0	0		
7	Diamond A	1995	0		1	9	0	11	0	0	0	0				12			
7	Diamond A	1997	0		6	2	0	2	0	1	1	0				39			
7	Cochran	1996	2		11	27	4	183	0	0	6	1				44			
7	Cochran	1997	0		25	3	0	52	0	2	13	0				31			
7	Cochran	1998	0		24	4	1	13	1	6	130	0				9			
7	Box O Wash	1995	68		0	1	0	152	0	0	0	0				28			
7	Box O Wash	1997	0		5	5	1	5	0	47	0	3				2			
7	Box O Wash	1998	0		39	17	5	38	1	103	1	1				10			

Appendix 2. Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
AGCH	–0.10	1.00	0.91	0.98	1.00	0.88	1.00	1.00	1.00	0.65	1.00	1.00	0.22
AGCH	–0.09	1.00	0.90	0.97	1.00	0.87	1.00	1.00	1.00	0.61	1.00	1.00	0.19
AGCH	–0.08	0.99	0.85	0.95	0.99	0.83	1.00	0.99	1.00	0.54	1.00	1.00	0.15
AGCH	–0.07	0.97	0.78	0.92	0.99	0.79	0.99	0.98	1.00	0.48	1.00	1.00	0.16
AGCH	–0.06	0.95	0.72	0.89	0.98	0.70	0.99	0.96	0.98	0.40	1.00	1.00	0.11
AGCH	–0.05	0.92	0.67	0.84	0.95	0.64	0.97	0.90	0.96	0.33	1.00	1.00	0.10
AGCH	–0.04	0.82	0.55	0.70	0.89	0.54	0.93	0.83	0.92	0.24	1.00	1.00	0.09
AGCH	–0.03	0.70	0.40	0.56	0.77	0.39	0.82	0.69	0.82	0.17	0.97	1.00	0.08
AGCH	–0.02	0.48	0.23	0.37	0.55	0.25	0.58	0.48	0.62	0.10	0.87	1.00	0.06
AGCH	–0.01	0.19	0.09	0.16	0.25	0.11	0.25	0.28	0.38	0.06	0.54	1.00	0.05
AGCH	0.00	0.03	0.02	0.01	0.02	0.04	0.01	0.08	0.22	0.02	0.02	0.05	0.05
AGCH	0.01	0.19	0.09	0.17	0.26	0.12	0.26	0.28	0.40	0.04	0.61	1.00	0.06
AGCH	0.02	0.52	0.26	0.41	0.58	0.26	0.64	0.53	0.67	0.10	0.95	1.00	0.06
AGCH	0.03	0.77	0.46	0.62	0.85	0.45	0.85	0.76	0.86	0.17	1.00	1.00	0.09
AGCH	0.04	0.90	0.64	0.82	0.95	0.56	0.97	0.88	0.96	0.24	1.00	1.00	0.14
AGCH	0.05	0.96	0.75	0.91	0.98	0.71	1.00	0.97	0.99	0.34	1.00	1.00	0.19
AGCH	0.06	1.00	0.86	0.97	1.00	0.82	1.00	0.99	1.00	0.43	1.00	1.00	0.23
AGCH	0.07	1.00	0.92	0.99	1.00	0.90	1.00	0.99	1.00	0.52	1.00	1.00	0.35
AGCH	0.08	1.00	0.97	1.00	1.00	0.94	1.00	1.00	1.00	0.58	1.00	1.00	0.43
AGCH	0.09	1.00	0.99	1.00	1.00	0.98	1.00	1.00	1.00	0.69	1.00	1.00	0.50
AGCH	0.10	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	0.72	1.00	1.00	0.60
AMME	–0.10	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00		
AMME	–0.09	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00		
AMME	–0.08	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.02	0.03	0.00		
AMME	–0.07	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.02	0.02	0.00		
AMME	–0.06	0.00	0.02	0.00	0.00	0.00	0.01	0.01	0.03	0.02	0.00		
AMME	–0.05	0.00	0.01	0.01	0.00	0.01	0.01	0.02	0.05	0.04	0.00		
AMME	–0.04	0.01	0.02	0.02	0.01	0.01	0.03	0.03	0.09	0.05	0.00		
AMME	–0.03	0.03	0.02	0.02	0.02	0.01	0.04	0.05	0.12	0.06	0.00		

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
AMME	–0.02	0.02	0.02	0.05	0.03	0.02	0.08	0.07	0.12	0.08	0.01		
AMME	–0.01	0.04	0.01	0.01	0.05	0.02	0.04	0.04	0.09	0.10	0.02		
AMME	0.00	0.02	0.01	0.01	0.03	0.01	0.02	0.04	0.08	0.01	0.04		
AMME	0.01	0.03	0.01	0.03	0.06	0.01	0.05	0.05	0.12	0.08	0.03		
AMME	0.02	0.03	0.02	0.04	0.04	0.01	0.07	0.09	0.14	0.05	0.02		
AMME	0.03	0.03	0.03	0.03	0.02	0.03	0.05	0.06	0.15	0.04	0.00		
AMME	0.04	0.02	0.04	0.02	0.01	0.02	0.05	0.05	0.11	0.02	0.00		
AMME	0.05	0.00	0.03	0.01	0.01	0.01	0.04	0.04	0.08	0.02	0.00		
AMME	0.06	0.00	0.02	0.00	0.00	0.01	0.03	0.03	0.05	0.03	0.00		
AMME	0.07	0.00	0.02	0.00	0.00	0.01	0.02	0.01	0.03	0.02	0.00		
AMME	0.08	0.00	0.02	0.00	0.00	0.00	0.01	0.01	0.03	0.03	0.00		
AMME	0.09	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.03	0.04	0.00		
AMME	0.10	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.00		
AMNA	–0.10	0.99	0.92	0.98	1.00	0.82	0.99	1.00	1.00	0.66	1.00	1.00	0.20
AMNA	–0.09	0.99	0.86	0.97	1.00	0.80	0.99	1.00	1.00	0.59	1.00	1.00	0.18
AMNA	–0.08	0.97	0.84	0.94	0.99	0.76	0.97	0.99	1.00	0.57	1.00	1.00	0.17
AMNA	–0.07	0.96	0.79	0.91	0.99	0.71	0.99	0.98	0.99	0.47	1.00	1.00	0.16
AMNA	–0.06	0.93	0.73	0.85	0.96	0.66	0.95	0.96	0.99	0.42	1.00	1.00	0.15
AMNA	–0.05	0.88	0.65	0.81	0.92	0.56	0.89	0.91	0.97	0.35	1.00	1.00	0.13
AMNA	–0.04	0.80	0.53	0.70	0.86	0.52	0.83	0.84	0.94	0.25	0.99	1.00	0.11
AMNA	–0.03	0.68	0.40	0.57	0.74	0.37	0.72	0.69	0.84	0.17	0.95	1.00	0.12
AMNA	–0.02	0.45	0.27	0.40	0.54	0.22	0.55	0.55	0.69	0.11	0.85	1.00	0.09
AMNA	–0.01	0.23	0.10	0.17	0.25	0.10	0.26	0.28	0.39	0.05	0.52	0.93	0.07
AMNA	0.00	0.04	0.03	0.03	0.05	0.04	0.02	0.09	0.20	0.03	0.04	0.05	0.04
AMNA	0.01	0.24	0.11	0.18	0.27	0.09	0.25	0.30	0.39	0.04	0.53	0.96	0.07
AMNA	0.02	0.47	0.28	0.41	0.53	0.26	0.57	0.54	0.67	0.12	0.88	1.00	0.07
AMNA	0.03	0.69	0.45	0.60	0.76	0.43	0.76	0.72	0.87	0.20	0.98	1.00	0.07
AMNA	0.04	0.79	0.58	0.72	0.88	0.59	0.89	0.84	0.94	0.30	1.00	1.00	0.06
AMNA	0.05	0.91	0.71	0.84	0.94	0.72	0.96	0.92	0.98	0.36	1.00	1.00	0.06

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
AMNA	0.06	0.95	0.80	0.92	0.98	0.82	0.98	0.96	1.00	0.43	1.00	1.00	0.08
AMNA	0.07	0.98	0.84	0.94	0.99	0.87	1.00	0.99	1.00	0.49	1.00	1.00	0.10
AMNA	0.08	0.99	0.91	0.97	1.00	0.93	1.00	0.99	1.00	0.54	1.00	1.00	0.13
AMNA	0.09	1.00	0.95	0.98	1.00	0.95	1.00	1.00	1.00	0.61	1.00	1.00	0.17
AMNA	0.10	1.00	0.96	1.00	1.00	0.98	1.00	1.00	1.00	0.64	1.00	1.00	0.20
CAIN	–0.10	1.00	0.98	1.00	1.00	0.96	1.00	1.00	1.00	0.86	1.00		
CAIN	–0.09	1.00	0.97	1.00	1.00	0.96	1.00	1.00	1.00	0.83	1.00		
CAIN	–0.08	1.00	0.96	1.00	1.00	0.93	1.00	1.00	1.00	0.75	1.00		
CAIN	–0.07	0.99	0.92	0.98	1.00	0.90	1.00	1.00	1.00	0.73	1.00		
CAIN	–0.06	0.99	0.90	0.97	1.00	0.88	1.00	1.00	1.00	0.63	1.00		
CAIN	–0.05	0.98	0.83	0.95	0.99	0.79	1.00	0.99	1.00	0.58	1.00		
CAIN	–0.04	0.95	0.76	0.89	0.98	0.75	1.00	0.97	0.99	0.46	1.00		
CAIN	–0.03	0.89	0.67	0.81	0.94	0.57	0.99	0.91	0.96	0.36	1.00		
CAIN	–0.02	0.74	0.48	0.64	0.81	0.36	0.93	0.78	0.88	0.19	0.97		
CAIN	–0.01	0.43	0.15	0.31	0.47	0.13	0.62	0.46	0.58	0.07	0.80		
CAIN	0.00	0.02	0.02	0.02	0.02	0.02	0.02	0.06	0.21	0.01	0.02		
CAIN	0.01	0.41	0.21	0.32	0.51	0.15	0.65	0.46	0.62	0.07	0.90		
CAIN	0.02	0.83	0.52	0.73	0.89	0.41	0.97	0.83	0.92	0.22	1.00		
CAIN	0.03	0.98	0.76	0.92	0.98	0.66	1.00	0.97	0.99	0.38	1.00		
CAIN	0.04	1.00	0.88	0.98	1.00	0.80	1.00	1.00	1.00	0.50	1.00		
CAIN	0.05	1.00	0.97	1.00	1.00	0.89	1.00	1.00	1.00	0.64	1.00		
CAIN	0.06	1.00	0.99	1.00	1.00	0.95	1.00	1.00	1.00	0.75	1.00		
CAIN	0.07	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.83	1.00		
CAIN	0.08	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.88	1.00		
CAIN	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00		
CAIN	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00		
CYCA	–0.10	0.95	0.82	0.93	0.97	0.58	0.95	0.99	1.00	0.61	0.99		
CYCA	–0.09	0.95	0.80	0.89	0.97	0.56	0.93	0.98	1.00	0.54	0.99		
CYCA	–0.08	0.93	0.74	0.86	0.96	0.54	0.92	0.97	1.00	0.49	0.99		

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
CYCA	–0.07	0.89	0.74	0.84	0.92	0.50	0.89	0.96	0.99	0.42	0.99		
CYCA	–0.06	0.85	0.68	0.78	0.89	0.46	0.84	0.93	0.98	0.37	0.99		
CYCA	–0.05	0.77	0.57	0.71	0.84	0.41	0.78	0.87	0.96	0.30	0.98		
CYCA	–0.04	0.70	0.49	0.62	0.72	0.37	0.75	0.78	0.93	0.22	0.95		
CYCA	–0.03	0.61	0.37	0.49	0.65	0.27	0.66	0.69	0.83	0.15	0.88		
CYCA	–0.02	0.42	0.22	0.35	0.49	0.16	0.52	0.46	0.68	0.08	0.75		
CYCA	–0.01	0.22	0.09	0.15	0.27	0.07	0.30	0.26	0.42	0.03	0.49		
CYCA	0.00	0.03	0.02	0.04	0.02	0.02	0.02	0.09	0.18	0.02	0.02		
CYCA	0.01	0.23	0.07	0.16	0.28	0.08	0.28	0.25	0.46	0.06	0.50		
CYCA	0.02	0.47	0.23	0.40	0.47	0.18	0.54	0.53	0.69	0.10	0.74		
CYCA	0.03	0.58	0.41	0.53	0.63	0.39	0.70	0.68	0.81	0.19	0.91		
CYCA	0.04	0.69	0.51	0.62	0.74	0.50	0.79	0.80	0.92	0.23	0.96		
CYCA	0.05	0.80	0.61	0.74	0.82	0.64	0.84	0.87	0.96	0.33	0.99		
CYCA	0.06	0.87	0.73	0.81	0.90	0.73	0.90	0.92	0.99	0.43	1.00		
CYCA	0.07	0.90	0.76	0.84	0.92	0.81	0.94	0.95	0.99	0.47	1.00		
CYCA	0.08	0.94	0.81	0.89	0.96	0.86	0.97	0.98	1.00	0.52	1.00		
CYCA	0.09	0.97	0.87	0.93	0.98	0.89	0.99	0.99	1.00	0.56	1.00		
CYCA	0.10	0.97	0.87	0.95	0.99	0.91	0.99	0.99	1.00	0.62	1.00		
CYLU	–0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	–0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	–0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	–0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	–0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00		
CYLU	–0.05	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.95	1.00		
CYLU	–0.04	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.86	1.00		
CYLU	–0.03	1.00	0.97	1.00	1.00	0.88	1.00	1.00	1.00	0.73	1.00		
CYLU	–0.02	0.99	0.89	0.97	1.00	0.73	1.00	1.00	1.00	0.56	1.00		
CYLU	–0.01	0.84	0.52	0.73	0.89	0.38	0.98	0.84	0.93	0.18	1.00		
CYLU	0.00	0.02	0.03	0.01	0.02	0.02	0.02	0.07	0.20	0.02	0.02		

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
CYLU	0.01	0.86	0.56	0.77	0.92	0.41	0.99	0.86	0.95	0.22	1.00		
CYLU	0.02	1.00	0.94	0.99	1.00	0.79	1.00	1.00	1.00	0.55	1.00		
CYLU	0.03	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.80	1.00		
CYLU	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00		
CYLU	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00		
CYLU	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
CYLU	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
GAAF	–0.10	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.87	1.00	1.00	0.63
GAAF	–0.09	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.82	1.00	1.00	0.59
GAAF	–0.08	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	0.77	1.00	1.00	0.51
GAAF	–0.07	1.00	1.00	1.00	1.00	0.93	1.00	1.00	1.00	0.67	1.00	1.00	0.44
GAAF	–0.06	1.00	0.99	0.98	1.00	0.86	1.00	1.00	1.00	0.59	1.00	1.00	0.35
GAAF	–0.05	0.99	0.97	0.97	1.00	0.79	1.00	0.98	1.00	0.48	1.00	1.00	0.25
GAAF	–0.04	0.98	0.92	0.93	0.99	0.66	1.00	0.97	0.99	0.37	1.00	1.00	0.20
GAAF	–0.03	0.90	0.81	0.80	0.94	0.51	0.98	0.90	0.95	0.28	1.00	1.00	0.11
GAAF	–0.02	0.70	0.57	0.55	0.78	0.32	0.88	0.69	0.79	0.14	0.98	1.00	0.11
GAAF	–0.01	0.28	0.22	0.23	0.42	0.12	0.52	0.37	0.50	0.05	0.77	1.00	0.06
GAAF	0.00	0.03	0.03	0.02	0.02	0.04	0.02	0.09	0.21	0.03	0.03	0.05	0.06
GAAF	0.01	0.32	0.25	0.25	0.42	0.14	0.51	0.36	0.51	0.05	0.82	1.00	0.06
GAAF	0.02	0.76	0.64	0.62	0.84	0.34	0.93	0.75	0.85	0.15	1.00	1.00	0.08
GAAF	0.03	0.93	0.87	0.84	0.97	0.59	1.00	0.93	0.97	0.27	1.00	1.00	0.14
GAAF	0.04	0.99	0.97	0.96	1.00	0.78	1.00	0.99	1.00	0.41	1.00	1.00	0.24
GAAF	0.05	1.00	0.99	0.99	1.00	0.89	1.00	1.00	1.00	0.54	1.00	1.00	0.36
GAAF	0.06	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.65	1.00	1.00	0.49
GAAF	0.07	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.75	1.00	1.00	0.65
GAAF	0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83	1.00	1.00	0.78

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
GAAF	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87	1.00	1.00	0.88
GAAF	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	0.94
ICPU	–0.10	0.96	0.78	0.91	0.98	0.74	0.98	0.98	1.00	0.53	1.00		
ICPU	–0.09	0.94	0.74	0.89	0.96	0.70	0.97	0.97	1.00	0.49	1.00		
ICPU	–0.08	0.90	0.72	0.85	0.93	0.63	0.95	0.95	1.00	0.43	1.00		
ICPU	–0.07	0.88	0.65	0.78	0.91	0.58	0.94	0.91	0.99	0.36	1.00		
ICPU	–0.06	0.84	0.59	0.74	0.88	0.47	0.90	0.87	0.97	0.30	0.99		
ICPU	–0.05	0.76	0.51	0.69	0.82	0.39	0.88	0.81	0.93	0.25	0.98		
ICPU	–0.04	0.68	0.43	0.60	0.75	0.30	0.83	0.75	0.87	0.17	0.93		
ICPU	–0.03	0.56	0.30	0.45	0.61	0.20	0.72	0.63	0.78	0.13	0.87		
ICPU	–0.02	0.40	0.16	0.30	0.47	0.13	0.55	0.44	0.61	0.07	0.75		
ICPU	–0.01	0.16	0.07	0.12	0.20	0.06	0.20	0.22	0.38	0.04	0.46		
ICPU	0.00	0.04	0.04	0.05	0.03	0.04	0.02	0.10	0.20	0.03	0.03		
ICPU	0.01	0.15	0.07	0.13	0.22	0.08	0.23	0.23	0.38	0.04	0.55		
ICPU	0.02	0.45	0.23	0.37	0.51	0.14	0.59	0.49	0.66	0.08	0.83		
ICPU	0.03	0.67	0.38	0.57	0.72	0.24	0.79	0.72	0.83	0.13	0.95		
ICPU	0.04	0.77	0.57	0.71	0.82	0.41	0.91	0.83	0.92	0.21	0.99		
ICPU	0.05	0.85	0.68	0.80	0.91	0.54	0.95	0.89	0.96	0.30	1.00		
ICPU	0.06	0.91	0.76	0.85	0.93	0.66	0.99	0.94	0.98	0.36	1.00		
ICPU	0.07	0.95	0.81	0.92	0.98	0.73	1.00	0.97	0.99	0.47	1.00		
ICPU	0.08	0.97	0.88	0.93	0.99	0.82	1.00	0.99	1.00	0.54	1.00		
ICPU	0.09	0.99	0.93	0.97	1.00	0.87	1.00	0.99	1.00	0.58	1.00		
ICPU	0.10	1.00	0.93	0.99	1.00	0.89	1.00	1.00	1.00	0.66	1.00		
LECY	–0.10	0.94	0.72	0.85	0.96	0.75	1.00	0.98	1.00	0.53	1.00	1.00	0.14
LECY	–0.09	0.91	0.70	0.82	0.94	0.69	0.99	0.96	0.99	0.53	1.00	1.00	0.15
LECY	–0.08	0.88	0.66	0.78	0.90	0.62	0.99	0.93	0.99	0.45	1.00	1.00	0.13
LECY	–0.07	0.84	0.65	0.76	0.89	0.56	0.97	0.91	0.98	0.40	0.99	1.00	0.12
LECY	–0.06	0.81	0.57	0.73	0.86	0.47	0.94	0.87	0.97	0.35	0.98	1.00	0.09
LECY	–0.05	0.80	0.57	0.67	0.80	0.43	0.93	0.82	0.93	0.29	0.94	1.00	0.07

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
LECY	–0.04	0.71	0.45	0.60	0.75	0.36	0.82	0.77	0.88	0.19	0.91	1.00	0.07
LECY	–0.03	0.59	0.35	0.49	0.63	0.28	0.69	0.64	0.80	0.16	0.87	1.00	0.07
LECY	–0.02	0.38	0.20	0.32	0.50	0.14	0.51	0.45	0.59	0.11	0.75	0.92	0.07
LECY	–0.01	0.16	0.08	0.14	0.22	0.07	0.24	0.24	0.37	0.05	0.47	0.49	0.06
LECY	0.00	0.04	0.03	0.04	0.05	0.05	0.02	0.09	0.20	0.04	0.04	0.05	0.05
LECY	0.01	0.18	0.07	0.12	0.20	0.08	0.26	0.21	0.35	0.05	0.50	0.48	0.05
LECY	0.02	0.40	0.21	0.34	0.47	0.14	0.55	0.47	0.62	0.09	0.85	0.98	0.06
LECY	0.03	0.62	0.35	0.53	0.67	0.28	0.73	0.65	0.77	0.13	0.99	1.00	0.05
LECY	0.04	0.74	0.51	0.67	0.84	0.37	0.87	0.81	0.90	0.23	1.00	1.00	0.07
LECY	0.05	0.89	0.67	0.81	0.94	0.43	0.97	0.91	0.96	0.27	1.00	1.00	0.11
LECY	0.06	0.96	0.75	0.89	0.98	0.59	0.99	0.97	0.99	0.36	1.00	1.00	0.16
LECY	0.07	0.98	0.83	0.96	0.99	0.68	0.99	0.98	1.00	0.44	1.00	1.00	0.22
LECY	0.08	1.00	0.91	0.98	1.00	0.76	1.00	1.00	1.00	0.48	1.00	1.00	0.26
LECY	0.09	1.00	0.94	0.99	1.00	0.85	1.00	1.00	1.00	0.51	1.00	1.00	0.37
LECY	0.10	1.00	0.97	1.00	1.00	0.90	1.00	1.00	1.00	0.61	1.00	1.00	0.50
MISA	–0.10	0.20	0.21	0.23	0.19	0.03	0.69	0.60	0.89	0.21	0.01		
MISA	–0.09	0.24	0.21	0.21	0.26	0.04	0.69	0.60	0.87	0.17	0.01		
MISA	–0.08	0.29	0.19	0.26	0.26	0.06	0.66	0.59	0.87	0.16	0.03		
MISA	–0.07	0.31	0.19	0.27	0.31	0.06	0.64	0.57	0.85	0.14	0.07		
MISA	–0.06	0.34	0.18	0.30	0.34	0.06	0.61	0.53	0.80	0.15	0.13		
MISA	–0.05	0.33	0.19	0.30	0.36	0.08	0.57	0.53	0.73	0.12	0.18		
MISA	–0.04	0.35	0.18	0.28	0.33	0.07	0.50	0.47	0.68	0.08	0.27		
MISA	–0.03	0.29	0.12	0.26	0.31	0.08	0.41	0.38	0.59	0.05	0.32		
MISA	–0.02	0.23	0.08	0.17	0.26	0.05	0.28	0.31	0.46	0.03	0.34		
MISA	–0.01	0.11	0.06	0.06	0.11	0.04	0.14	0.16	0.29	0.02	0.25		
MISA	0.00	0.02	0.02	0.02	0.02	0.02	0.01	0.05	0.18	0.01	0.03		
MISA	0.01	0.10	0.04	0.06	0.11	0.05	0.13	0.16	0.25	0.02	0.27		
MISA	0.02	0.23	0.10	0.19	0.29	0.09	0.33	0.34	0.48	0.03	0.52		
MISA	0.03	0.36	0.20	0.29	0.40	0.14	0.48	0.46	0.64	0.05	0.66		

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
MISA	0.04	0.47	0.29	0.40	0.51	0.15	0.64	0.55	0.75	0.08	0.79		
MISA	0.05	0.54	0.34	0.50	0.57	0.19	0.80	0.67	0.86	0.13	0.89		
MISA	0.06	0.59	0.45	0.54	0.64	0.20	0.87	0.77	0.90	0.15	0.97		
MISA	0.07	0.68	0.53	0.61	0.71	0.23	0.92	0.82	0.94	0.21	1.00		
MISA	0.08	0.72	0.50	0.69	0.79	0.27	0.97	0.87	0.97	0.25	1.00		
MISA	0.09	0.78	0.59	0.71	0.82	0.29	0.98	0.88	0.98	0.28	1.00		
MISA	0.10	0.79	0.65	0.75	0.87	0.32	0.99	0.92	0.99	0.35	1.00		
PACL	–0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.56
PACL	–0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.55
PACL	–0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	0.54
PACL	–0.07	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	0.54
PACL	–0.06	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	0.91	1.00	1.00	0.54
PACL	–0.05	1.00	0.98	0.99	1.00	0.98	0.99	1.00	1.00	0.86	1.00	1.00	0.51
PACL	–0.04	0.99	0.96	0.99	1.00	0.95	0.98	1.00	1.00	0.79	1.00	1.00	0.47
PACL	–0.03	0.98	0.93	0.97	0.99	0.86	0.96	0.99	0.99	0.63	1.00	1.00	0.41
PACL	–0.02	0.94	0.80	0.90	0.97	0.67	0.90	0.95	0.99	0.38	1.00	1.00	0.41
PACL	–0.01	0.69	0.39	0.59	0.74	0.25	0.72	0.70	0.85	0.16	0.97	0.99	0.42
PACL	0.00	0.03	0.03	0.04	0.03	0.03	0.02	0.10	0.21	0.02	0.03	0.07	0.01
PACL	0.01	0.70	0.42	0.62	0.77	0.31	0.73	0.74	0.84	0.15	0.97	1.00	0.39
PACL	0.02	0.95	0.83	0.90	0.96	0.75	0.89	0.96	0.98	0.42	1.00	1.00	0.38
PACL	0.03	0.98	0.93	0.97	0.99	0.93	0.93	0.99	1.00	0.66	1.00	1.00	0.33
PACL	0.04	1.00	0.96	0.98	1.00	0.98	0.97	0.99	1.00	0.81	1.00	1.00	0.31
PACL	0.05	1.00	0.98	0.99	1.00	0.99	0.98	1.00	1.00	0.90	1.00	1.00	0.29
PACL	0.06	1.00	0.98	1.00	1.00	1.00	0.99	1.00	1.00	0.92	1.00	1.00	0.28
PACL	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.24
PACL	0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	0.24
PACL	0.09	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.25
PACL	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.24
PIPR	–0.10	0.10	0.10	0.09	0.07	0.03	0.55	0.33	0.65	0.10	0.00		

Appendix 2 (continued). Power of different sampling scenarios for detecting trends. Scenario abbreviations as in Table 2.

		Sampling scenario											
species	trend	1	2	3	4	5	6	7	8	9	10	13	14
PIPR	–0.09	0.11	0.10	0.14	0.11	0.04	0.54	0.31	0.67	0.09	0.01		
PIPR	–0.08	0.13	0.12	0.15	0.12	0.04	0.52	0.35	0.59	0.10	0.01		
PIPR	–0.07	0.15	0.11	0.13	0.17	0.04	0.47	0.33	0.58	0.09	0.04		
PIPR	–0.06	0.17	0.10	0.14	0.17	0.05	0.43	0.29	0.52	0.07	0.06		
PIPR	–0.05	0.17	0.08	0.15	0.18	0.04	0.41	0.27	0.51	0.06	0.08		
PIPR	–0.04	0.15	0.08	0.13	0.18	0.03	0.32	0.26	0.45	0.05	0.13		
PIPR	–0.03	0.12	0.07	0.10	0.15	0.04	0.23	0.22	0.36	0.03	0.18		
PIPR	–0.02	0.07	0.05	0.08	0.10	0.04	0.17	0.16	0.29	0.03	0.18		
PIPR	–0.01	0.05	0.03	0.04	0.05	0.03	0.06	0.10	0.22	0.02	0.11		
PIPR	0.00	0.04	0.03	0.03	0.05	0.03	0.03	0.09	0.21	0.03	0.04		
PIPR	0.01	0.06	0.03	0.05	0.06	0.03	0.08	0.11	0.22	0.04	0.14		
PIPR	0.02	0.12	0.05	0.10	0.13	0.05	0.19	0.18	0.33	0.04	0.28		
PIPR	0.03	0.17	0.10	0.17	0.21	0.06	0.28	0.25	0.42	0.05	0.47		
PIPR	0.04	0.25	0.14	0.21	0.28	0.06	0.41	0.37	0.57	0.06	0.69		
PIPR	0.05	0.32	0.19	0.28	0.38	0.09	0.58	0.45	0.65	0.07	0.84		
PIPR	0.06	0.40	0.26	0.34	0.45	0.12	0.73	0.56	0.77	0.09	0.95		
PIPR	0.07	0.50	0.31	0.42	0.56	0.12	0.82	0.64	0.84	0.11	0.99		
PIPR	0.08	0.60	0.39	0.51	0.64	0.14	0.93	0.74	0.91	0.14	1.00		
PIPR	0.09	0.67	0.43	0.57	0.71	0.15	0.95	0.81	0.94	0.17	1.00		
PIPR	0.10	0.72	0.51	0.66	0.80	0.19	0.98	0.86	0.95	0.18	1.00		