Sampling Frequency for Quality Water Data, Thinking About the Sampling Frequency

Ronald F. Turco, rturco@purdue.edu, Megan Heller Haas, mheller@purdue.edu
College of Agriculture – Crops, Soils, and Environmental Sciences, Lilly Hall of Life Sciences,
Purdue University, West Lafayette, Indiana, USA

ABSTRACT

As an increasing population intensifies demands on the world's water supplies' questions are often raised about the best sampling practice to detect changes in water quality. However, it is often difficult to define the appropriate number of water samples to take within a given water-monitoring program. Therefore, we present a discussion on how to best define the number of samples required to assess changes in a watershed. A better defined plan will allow watershed managers a means to optimize their approaches to water quality protection.

Keywords: Water, sampling, quality, management.

1. INTRODUCTION

Of the many questions asked by watershed managers across the world, one common inquiry is: "As part of our watershed project how many samples do we need to collect and examine to understand what is going on with the water?" This question holds regardless of where the work is conducted, and the answer will impact the size, scope, and cost of the effort. The secondary version of the question is how many water samples do we need collect to detect a change caused by a new management or conservation practice? At the outset of considerations, the watershed manager should be aware that the number of water samples that will be needed will reflect the intrinsic variance in water systems studied. This point is often overlooked for a more idealistic approach of "needing to know what is out in the watershed." In order to show a significant difference from a management change, be prepared to collect more samples where the variance is high, as in streams and rivers, than where the variance is low, lakes, and ponds. There is not a right or wrong answer on sampling intensity; there are choices and consequences about what can be said from the findings. The most important consideration is "what do we want to know about our system" – this will govern the best approach. To this end, one should resolve a number of early issues including: (1) what materials are we interested in evaluating; (2) what level of change are we interested in detecting; (3) can we manage the sample collection analysis process; (4) are we interested in only concentration or do we want to measure flow volumes and consider the load of the materials in the water system; and the ultimate concern, (5) what level of confidence do we want to have in our findings (e.g., what do the numbers really mean and are they defensible at some level of statistical significance?) There are a number of statistical approaches that can be used to provide directions and inform this process.

To inform the process, the development of a sampling plan should include the collection of a preliminary set of samples – establishing a mean, standard deviation, and variance. The preliminary sampling should cover a significant length of time and include the types of weather conditions common in the watershed. Collecting five samples in 5 days is far different from collecting five samples ~5 months. An estimation of concentration mean and standard deviation will allow for a better prediction of initial levels of precision for the project. If resources are limited, utilize other studies from the region to inform the local process. Data presented later in this section give an indication of the types of variance you will typically encounter in both large and small Midwest streams. However, many aspects of water contamination are correlated. So, sometimes it is difficult to get an absolute answer. Another subtle issue to consider is that installing field scale implementation practices along a stream will only have an effect on that water coming into contact with the practice. The anticipated reductions in contamination for the overall watershed may be difficult to detect if practices only cover a small portion of the watershed.

Because of resource limitations, what occurs most often is the use of a fixed sampling plan (often quarterly) that is imposed on the project, and a consideration of data precision and the level of detectable difference is never actually stated. Statistically, using a fixed approach will limit the sensitivity of analysis and the degree of system change that is detectable (as the response is controlled by the variance in the system). Therefore, using a general fixed sampling plan will give a "view" of the system but is insensitive to the level of variance present. For example, if your goal is a 25% reduction in the level of nitrate, phosphorus, or *Escherichia coli* in a water system, your sampling strategy should be designed to indicate this level of reduction at a stated confidence level so you could detect if you have had an impact from a conservation

practice on the water quality in a given location. One could also use the approach to detect differences in two streams or in portions of a stream. The most important consideration is at what level of confidence (what level of significance) the findings can be held. With a fixed sampling protocol, the levels of "significant reduction" that is detectable and can be stated as significant (regardless of what you actually hope to achieve) are controlled by the imposed sampling plan (sample number) set inside the systems' level of variance.

This is fine except if ask the general question "did the imposed change significantly impact water quality"? For a watershed manager, a well-crafted data collection and analysis will clearly demonstrate if a management system has caused a significant difference. To be clear, when used in this context the word "significant" means the likelihood that an event could occur above a level controlled by only chance. So when we say the event is significant (e.g., a reduction in a nutrient level) at an alpha of 5%, it means the event occurs randomly 5% or less of the time, and we are 95% confident that the result is real. More importantly, by setting a desired level of significance, we can test to see if one data set (result 1) is different from another data set (result 2), and answer the question about the significance of the management change has made. This leads to a second point; to show you have had an impact on some aspect of water quality after you impose a practice, it is best to have collected data on quality prior to that practice being applied. If sampled correctly, this can provide irrefutable proof of the change. Moreover, even when the sampling frequency is less than ideal, a before and after approach can at least demonstrate numeric differences and aid in the discussion of the practice.

In using this approach, we are looking for help in establishing the number of samples needed to see a change in average values over a period of time with everything occurring in some level of sampling noise.

2. DETERMINING SAMPLE NUMBER

So the question becomes what is the "real signal and within the noise" and can we see a change in the signal? If we collect ~100 river samples and find a mean value as 0.148 mg of X L^{-1} with a standard deviation of 0.131 mg L^{-1} and then impose a practice with the goal of reducing the mean value in the water to 0.14 mg L^{-1} (a 5% reduction) assuming that the standard deviation in the second set of samples remains the same, from a statistical point-of-view, how many samples do we need to collect to prove this change is significant? Typically sampling plans are not presented with this approach.

To illuminate these points, a data set from water samples collected on three Indiana streams and the midsection of the Wabash River was used to show the importance of sampling frequency on the precision of results. In this study, we collected data on a number of parameters, but this discussion will be constrain to a consideration of concentration of *E. coli*, total suspended soils, total phosphorus (TP), ammonia, and nitrate as these are often at the core of many sampling programs. The general characteristics of the water systems are found in Table 1. It is clear that we are dealing with two classes of stream based on flow rates (the Wabash and the other three streams). Another approach to sample would be done considering load (which requires an estimate of flow) and is beyond this discussion. In terms of watershed management, the values in Table 1 should be considered pretreatment values, as we have not imposed any sort of land managements on the watershed systems.

The mean values for the five sampling locations and each targeted parameter are presented in Tables 2–6. For the systems, the values show high standard deviations directly connected to extremes in their concentration range. For example, *E. coli* numbers can range from 0 to ~92,000 cfu 100 mL^{-1} on the same stream. Further observation of the data sets indicates that extreme values are commonly encountered in all of the measured parameters. Therefore, picking the right sample number comes down to how we evaluate of the changes in the sample mean within a range of values created by the standard deviation. This is the interaction that actually controls how we should address future assessments in the system.

Table 1. Water quantity and associated land use data.

System	Mean (x) (cfs)	Standard deviation (d)	High flow (cfs)	Low flow (cfs)	Total area acr	Total AG (%)	Total Dev (%)	Total For (%)
Wabash R. 1	7,837	9,053	58,646	1,056				
Wabash R. 2	8,147	9,411	60,967	1,098				
Little Pine	25.7	55	800	0.02	13,855	89.6	7.3	2.6
Little Wea	24.9	50	1,413	0.41	11,067	93.8	4.3	1.2
Elliott Ditch	19.6	58	1,748	0.3	11,451	47.2	48	1.8

cfs, cubic feet per second; AG, agriculture; Dev, developed; For, forest.

Table 2. Summary of water quality data – *E. coli*

System	Mean (*x*) mg/L	Standard deviation (*d*)	Min mg/L	Max mg/L	Range	Count
Wabash R. 1	334	591	10	2489	2479	126
Wabash R. 2	663	1669	9	15531	15522	128
Little Pine	1671	3081	33	24195	24162	137
Little Wea	547	1237	12	9803	9791	139
Elliott Ditch	2452	9422	17	92084	92066	140

Table 3. Summary of water quality data – *Total Suspended Solids (TSS)*

System	Mean (*x*) mg/L	Standard deviation (*d*)	Min mg/L	Max mg/L	Range	Count
Wabash R. 1	39	49	1.2	272	271	126
Wabash R. 2	49	58	3.2	304	301	127
Little Pine	23	34	1.2	261	260	138
Little Wea	15	41	0.4	352	351	137
Elliott Ditch	12	25	0	172	172	138

Table 4. Summary of water quality data – *Total Phosphorus (P)*

System	Mean (*x*) mg/L	Standard deviation (*d*)	Min mg/L	Max mg/L	Range	Count
Wabash R. 1	0.07	0.08	0	0.62	0.62	125
Wabash R. 2	0.10	0.07	0	0.51	0.51	124
Little Pine	0.15	0.13	0.02	0.89	0.88	136
Little Wea	0.051	0.12	0	0.73	0.73	136
Elliott Ditch	0.061	0.345	0	2.76	2.76	139

Table 5. Summary of water quality data – *Ammonium-N (NH_{4+})*

System	Mean (*x*) mg/L	Standard deviation (*d*)	Min mg/L	Max mg/L	Range	Count
Wabash R. 1	0.049	0.21	0	1.6	1.6	125
Wabash R. 2	0.045	0.15	0	1.3	1.3	126
Little Pine	0.04	0.12	0	0.89	0.89	137
Little Wea	0.09	0.64	0	7.36	7.36	137
Elliott Ditch	0.08	0.34	0	2.76	2.76	139

Table 6. Summary of water quality data – *Nitrogen-N ($NO_{3/2-}$)*

System	Mean (*x*) mg/L	Standard deviation (*d*)	Min mg/L	Max mg/L	Range	Count
Wabash R. 1	3.08	2.07	0	8.3	8.3	125
Wabash R. 2	2.79	2.03	0	8.4	8.4	126
Little Pine	6.42	3.99	0.03	20.87	20.84	137
Little Wea	4.45	2.81	0.02	14.2	14.1	137
Elliott Ditch	1.14	0.74	0	4.28	4.28	139

3. APPROACH

To answer the question posed in Section 1, Dunnette (1980) and others (National Water Quality Handbook [NWQH], 2003) provide an approach for estimating sample numbers for a watershed study. This approach is designed to determine the true mean with a specified level of accuracy. They suggest use of the equation

$$n = \frac{z^2\sigma^2}{L^2} \quad (1)$$

n is the number of samples one "should" take, z is the confidence coefficient (the same as a Student's t factor), σ is the sample variance (the standard deviation squared), and L is the desired difference from the mean. For data sets where a large number of samples are used to create the mean, the z factor can be set to an infinity value, simplifying calculations while giving a good indication of n. As one makes calculations, no changes with the z value is required as long as n remains large. As more refined measures are made, when sample numbers drop <100, a reconsideration of the z value (Student's t-test value) should be undertaken.

For an example, we evaluated data from Little Pine creek to find the number of samples needed for a determination of a 5% reduction in TP at a 95% confidence level, which would be 1,268, using the above-described equation.

$$n = \frac{(1.962)^2(0.131)^2}{(0.148 \times 0.05)^2} \quad (2)$$

At this value of n, we would be 95% confident that a difference (up or down) of >5% from the current mean level of TP in the Little Pine creek can be detected. It must be reiterated that this value would be for before the management change, and a total number of $2n$ or 2,536 samples would be required to see a change of >5% from the implementation of a conservation practice (NWQH, 2003). The reason n is so large is that we have a highly variable system or σ (relative to the sample mean). In this case, the coefficient of variation (CV = σ/mean) is 88%. If the system had a 50% reduction in σ, this would reduce the n value to 22. Therefore, if a practice did reduce the variation in mean, this could be accounted for determining future sample planning. The σ value reflects the variance in the system, so reductions are unlikely, but it is critical to point out the factors motivating the n value. Clearly, natural systems are "messy," highly variable and crosscorrelated. If we reduce the level of confidence, we are willing to accept to 90%, the number of samples required also falls to 893. On the other hand, the value of L can be adjusted to meet considerations of concentration reductions, for example, a desired level of reduction in TP. So if we are interested in observing for only large changes, say a 20% reduction from the mean (0.029 mg L^{-1}), then the number of samples required falls to 56 or 112 for a pre- and postimplementation study. In this approach, a small change in the mean (while numerically visible) will be nonsignificant in its difference.

Clearly, a major use of this approach is justified when sampling is tied to questions about the detection of reductions. For example, on Little Pine TP levels' average is 0.15 mg L^{-1}. To reduce this to a target level of 0.08 mg L^{-1} (a 53% reduction), we have just installed a new conservation practice that has radically reduced the TP level. How many samples do we need to collect to show that a significant reduction has occurred? Using the above-desicribed equation, we find 95% confidence in our number, and we need a n value of 11 over the year to confirm a reduction to 0.08 mg L^{-1}. However, the finding would only be significant at TP levels of 0.08 mg L^{-1} or lower. That is, smaller average changes would appear as nonsignificant following statistical analysis. One must determine the level of reduction you need to achieve and build a sampling protocol to reach this point if you want to show the reduction.

Sampling is expensive, and we are frequently required to fix the number of samples collected to a schedule unrelated to σ. For example, sampling of a watershed four times a year is often used. While this is an important and widely applied approach, an often unasked question is "what level of difference can we detect?" Using the same equation $n = z^2\sigma^2/L^2$ but solving for L (where L = mean × difference), $L^2 = z^2\sigma^2/n$, and resetting z for four sample times or three degrees of freedom, we can find our critical level of significant difference. Again taking the Little Pine watershed data for TP and assuming we want a 95% confidence level for our findings, quarterly sampling would limit detecting significant changes in the mean level of TP to changes in excess of 96% or ±0.142 mg L^{-1}. In other words, the quarterly sampling relegates us to an assessment that will only allow us to show significance with large changes in the average level of TP in the water. If we retain the quarterly sampling plan but lower the confidence level to 90%, we would be limited to detect significant changes in the mean level of TP in excess of 69% or ±0.103 mg L^{-1}.

4. SAMPLING NEEDS TO REACH TARGET GOALS

From the information provided earlier, we can determine the number of samples needed for statistically indicating a given goal level. The percentage reduction depends on the system and the water quality goals for the location. Data in Table 7 show the calculated n values needed to statistically test for

Table 7. Minimum frequency of sampling per year required to establish an achievement of indicator level at either the 95 and 90% confidence levels.

	E. coli			Total Phosphorus			Nitrogen $NO_{3/2}$-N		
	Mean[1]	Confidence Level		Mean[2]	Confidence Level		Mean[3]	Confidence Level	
		95%	90%		95%	90%		95%	90%
System	cfu/100 mL	n		mgL^{-1}	n		mgL^{-1}	n	
Wabash R. 1	334	104	73	0.07	BI[4]	BI	3.08	14	10
Wabash R. 2	663	59	42	0.1	21	15	2.79	17	12
Little Pine	1671	18	12	0.15	15	10	6.42	3	2
Little Wea	547	59	41	0.051	BI	BI	4.45	5	4
Elliott Ditch	2452	69	49	0.061	BI	BI	1.14	BI	BI

[1] Indicator goal *E. coli* = 231 cfu /100 mL
[2] Indicator goal P = 0.08 mgL^{-1}
[3] Indicator goal $NO_{3/2}$-N = 2 mgL^{-1}
[4] BI=Below indicator level no reduction required

attainment of the stated indicator goals. These results show the detection of significant, but small changes in highly variable systems require more sampling than is required for the detection of larger changes – assuming you reach the goal level and the variance is similar. In all cases except for nitrate-N, quarterly sampling would not be frequent enough to confirm a reduction to the target level.

5. FINAL CONSIDERATIONS

Accurate determination of contaminant concentration in highly variable systems is difficult. This difficulty reflects the fact that the concentrations of some materials can change over 500-fold (Tables 2–6) across a sampling season. On the other hand, statistical estimation methods for establishing sampling needs and defining the levels of detectable outcomes are possible but seems to have taken a backseat to the use of prescribed sampling plans. If required to use a fixed sampling plan, it is critical to anticipate the level of reduction that can be, from a statistical view point, attained and deemed significant. This brings us to a critical question: Are we simply watching the systems and hoping to see some improvements or do we want to critically assess the changes that are occurring as a way of understanding our watershed? As we expend money on improvement practices, we must be sensitive to assess their impact and think about the needed sampling intensity. Clearly, small changes in water quality are going to be difficult to detect and confirm statistically unless more rigorous sampling protocols are employed or the treatment lessens the variability of the sample mean. This is complicated by the fact that we are assuming a practice that will make an almost instantaneous change in the target levels. In reality, it may take years to see the full reduction. We are also assuming the responses that are not correlated with season, but we expect contaminants like nitrate may violate this assumption. So considerations of mean and variance within a season may be in order.

Systems employing low-sampling frequencies must manage their “impact expectations” and think about the speed at which the applied implementation practice will be maximal in its effectiveness. That is, while changes could be seen, the ability to rigorously test the significance of the finding is going to be limited except where extreme changes are found. Staged sampling where the practice is given time to mature in its effectiveness and most of the samples are collected after this point may provide a cost savings. With both small changes in sample mean and infrequent sampling, it may be difficult to say, with statistical certainty, we are improving water quality following the implementation of a management practice.

6. STEPS SUGGESTED FOR TESTING WATER QUALITY CHANGES FROM AN IMPLEMENTATION PROJECT

(1) Decide what you want to measure and how you are going to conduct the process. Typically the group will be trying to limit the impairment.

(2) Establish a plan and collect preliminary data (over an appropriate length of time) for the water to estimate mean concentration and standard deviation of the contaminant. (This step is frequently overlooked but is critical.)

(3) From the preliminary data set, determine the level of change that is needed to meet your goal and select the field protocol that will lead to this change. From the existing literature, estimate the

best possible reduction that could be achieved with that practice and estimate the time required to see the maximal impact from the practice. This step is frequently overlooked, but the selection of a practice based on the desired level of reduction is cost effective.

(4) The time step in step 3 controls both the pre- and postimplementation sampling frequencies. If a fixed sampling protocol is used, calculate the best possible reduction that is detectable and use this to estimate the sampling time. Some consideration of the cost of the practice and sensitivity of the measurement can be undertaken.

(5) Begin to collect preimplementation data at the frequency indicated in step 4. Test this long-term data set against the short-term preliminary data and readjust the sampling frequency.

(6) At an appropriate time, install the field protocol and monitor at a similar frequency as indicted in step 5.

REFERENCES

Dunnette, D. A. (1980). Sampling frequency optimization using a water quality index. *Journal of the Water Pollution Control Federation*, *52*, 2807–2811.

National Water Quality Handbook. (2003). *Sampling frequency and duration* (pp. 9-1–9-6). Retrieved from http://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/national/water/quality/?&cid=stelprdb1044782. [450–VI–NWQH, September 2003].

Dr. Ronald F. Turco Jr. earned two B.S. degrees from the University of Idaho: the first is in Soil Science (1979) and his second is in Microbiology (1980). He received his Ph.D. in Soil Microbiology from Washington State University in 1985. Dr. Turco joined the faculty of Purdue University in 1985 and is currently a Professor in College of Agriculture. Dr. Turco is the Director of the Indiana Water Resource Research Center (IWRRC.org) and the Purdue Water Community. Dr. Turco's research program is divided into three major areas of investigation: the environmental fate of pathogens, the environmental fate of organic chemicals, and efforts to understand how soil is impacted by biofuel production. As the Indiana Water Center Director, Dr. Turco is also working on a project to address the revitalization of the Wabash River corridor by proving expertise that will help minimize the occurrence of pathogenic bacteria and excess nutrients. Dr. Turco teaches classes on soil ecology, soil microbiology and biodegradation, and bioremediation. Dr. Turco works with a number of research programs across the Purdue Campus and country. He collaborates with faculty in Purdue's Colleges of Engineering and Science as well as working with the Purdue Climate Change Research Center and the Center for the Environment.

Megan Heller Haas graduated from Purdue University in 2005 with a Bachelor's of Science degree in Natural Resources and Environmental Science. In 2013, she graduated with a Master's of Science degree in Agronomy with an emphasis in water resources. Currently, Megan works in an Agronomy Research Lab at Purdue University. Past research projects have included degradation of biofuels in soil and methane production from waste materials. Since 2009, Megan has been investigating the water quality of local watersheds.