CHAPTER 1

NORMAL PREDICTION INTERVALS

COPYRIGHTED MATERIAL

1.1 OVERVIEW

The fundamental problem in groundwater monitoring is the prediction of future measurements based on a background sample of historical measurements [Davis and McNichols, 1987; Gibbons, 1987b]. In some cases, the background sample may consist of repeated measurements from a collection of wells located upgradient of the facility. In other cases, the background sample may consist of repeated measurements from a single monitoring well to which the consistency of future measurements will be compared (i.e., intra-well comparison). In either case, if the number of future comparisons is finite and known, we may wish to compute an interval that will contain all future measurements with a given level of confidence (e.g., 95% confidence). The most critical problem is correctly defining the number of future comparisons and constructing the corresponding statistical decision rule so that the confidence level pertains to the site as a whole. As will be shown, the number of future comparisons includes the total number of constituents and monitoring wells for which a statistical test is to be performed. To assume any less will result in countless site assessments and possible corrective action due to chance alone.

To provide a statistical foundation, denote the number of future comparisons as k and the confidence level $1 - \alpha$, where α represents the false positive rate or Type I error rate of the decision rule. The false positive rate is the rate at which we would reject a new value if in fact it came from the same distribution as the background measurements.

The appropriate statistical interval for this application is known as a prediction interval [Davis and McNichols, 1987; Gibbons, 1987b; Gibbons and Baker, 1991]. A synonym for prediction interval is a beta-expectation tolerance interval, in that on average, the new measurements will be contained with confidence level $1 - \alpha$.

In the context of groundwater monitoring, prediction intervals play an important role because we often know the number of statistical comparisons made on each monitoring event, and for regulatory purposes, we must include all measurements or risk a potentially costly site assessment. However, what constitutes the number of future measurements is not always trivial. Is it the number of monitoring wells, the number of constituents at a particular monitoring well or the combination of both? Should the number of future comparisons be restricted to those performed on the next single monitoring event or for all future monitoring events? To answer these questions, it is important to understand the consequences of a false positive decision and the impact the choice of k has on the false positive and negative rates of the statistical test. The false negative rate describes the frequency of failure to reject a new measurement when it has come from a different distribution than the background measurements. To better understand the answers to these questions, let us begin with the simplest form of prediction limit: a prediction limit for the next single measurement from a normal distribution.

1.2 PREDICTION INTERVALS FOR THE NEXT SINGLE MEASUREMENT FROM A NORMAL DISTRIBUTION

Assume we have collected $n = 8$ background water quality measurements for total organic carbon (TOC) levels, denoted $x_1, \ldots, x_8$. The sample mean and sample standard deviation of these eight measurements are given by

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{1.1}$$

and

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}. \tag{1.2}$$

On the next quarterly monitoring event, we intend to collect a new TOC measurement from the same well or a corresponding compliance well located downgradient of the facility. Based on the previous eight samples, what interval will contain the next single TOC measurement with $(1 - \alpha)100\%$ confidence?

To construct such an interval, we must begin by examining the sources of uncertainty in this problem. First, note that $\bar{x}$ and s are merely sample-based estimates of the true population mean and standard deviation μ and σ. If we had measured all groundwater in the area for that time period, $\bar{x}$ would equal μ and s would equal σ; however, we only have eight available measurements; hence, we will have considerable uncertainty in our estimates of μ and σ. Fortunately, we can quantify our uncertainty in σ by noting that the sample mean $\bar{x}$ is distributed normally with mean μ and standard deviation $\sigma/\sqrt{n}$ (i.e., $\bar{x} \sim N[\mu, \sigma^2/n]$). Second, note that the new measurement x_{new} also has an associated

measurement error σ for which we have a sample-based estimate s and is independent of the prior measurements. Combining these two sources of uncertainty and selecting the $(1-\alpha/2)100\%$ point of Student's t-distribution with $n-1$ degrees of freedom yields the interval

$$\bar{x} \pm t_{[n-1,1-\alpha/2]}\sqrt{s^2 + \frac{s^2}{n}}\,, \tag{1.3}$$

which can be expressed in the more familiar form

$$\bar{x} \pm t_{[n-1,1-\alpha/2]}s\sqrt{1 + \frac{1}{n}}\,. \tag{1.4}$$

This interval will provide $100(1-\alpha)\%$ confidence of including the next future measurement from the normal distribution for which we have a sample of n previous measurements.

Frequently, however, we are most interested in providing an upper limit for the new measurement, since, for example, a TOC measurement that is too low poses no environmental threat. In this case, we compute the one-sided normal prediction limit as

$$\bar{x} + t_{[n-1,1-\alpha]}s\sqrt{1 + \frac{1}{n}}\,. \tag{1.5}$$

This prediction limit provides $(1-\alpha)100\%$ confidence of not being exceeded by the next single measurement.

Example 1.1

Consider the data in Table 1.1 for TOC measurements from a single well over two years of quarterly monitoring.

Table 1.1: Eight Quarterly TOC Measurements

Year	Quarter	TOC in mg/L
1992	1	10.0
1992	2	11.5
1992	3	11.0
1992	4	10.6
1993	1	10.9
1993	2	12.0
1993	3	11.3
1993	4	10.7

Inspection of the data reveals no obvious trends, and these data have mean $\bar{x} = 11.0$ and standard deviation $s = 0.61$. The upper 95% point of Student's t distribution with $n-1 = 8-1 = 7$ degrees of freedom is $t_{[7,1-.05]} = 1.895$; therefore, the upper 95% confidence normal prediction limit is given by

$$11.0 + 1.895(0.61)\sqrt{1 + \frac{1}{8}} \;=\; 12.23 \text{ mg/L}\,,$$

which is larger than any of the observed values. Had we required 99% confidence of including the next single measurement, the upper 99% point of Student's t distribution on 7 degrees of freedom is $t_{[7,.01]} = 2.998$; therefore, the upper 99% confidence normal prediction limit is given by

$$11.0 + 2.998(0.61)\sqrt{1 + \frac{1}{8}} \;=\; 12.94 \text{ mg/L} .$$

These limits (i.e., 12.23 mg/L and 12.94 mg/L) provide 95% and 99% confidence respectively, of including the next single observation from a normal distribution for which eight previous measurements have been obtained with observed mean 11.0 mg/L and standard deviation 0.61 mg/L.

1.3 PREDICTION LIMITS FOR THE NEXT k MEASUREMENTS FROM A NORMAL DISTRIBUTION

In practice, it is rare to have an application in which only a single future measurement requires evaluation. Typically, TOC measurements are obtained from a series of downgradient or compliance wells and must be simultaneously evaluated. The simplest approach is to assume independence. Under independence, if the probability of a false positive result for a single comparison is α, the probability of at least one of k comparisons being significant by chance alone is

$$\alpha^* = 1 - (1 - \alpha)^{k^*} . \tag{1.6}$$

Here, α^* is the site-wide false positive rate since it simultaneously considers all k^* comparisons being performed on a given monitoring event. At this point, we consider k^* to represent the total number of comparisons which is the product of the number of monitoring wells and constituents ($k^* = kc$). For example, with 95% confidence for an individual comparison (i.e., $\alpha = .05$) and $k^* = 10$ comparisons, the probability of at least one significant result by chance alone is

$$\alpha^* = 1 - (1 - .05)^{10} = .40 ,$$

or a 40% chance of a statistically significant exceedance by chance alone. With 100 comparisons, $\alpha^* = .99$ or a 99% chance of a statistically significant exceedance by chance alone. Since it is not uncommon for detection monitoring programs to have 20 or 30 monitoring wells, each monitored quarterly for 10 or 20 constituents (in some cases far more), the effect of these multiple comparisons on the site-wide false positive rate is considerable. The likelihood of chance failure is near certainty. A facility with 25 wells, each monitored for 20 constituents, will be performing 500 statistical tests per sampling event. Even setting $\alpha = .01$ will produce a probability of $\alpha^* = .99$ or a 99% chance of failing at least one of those tests by chance alone. Since most state and federal regulations require costly site assessments that may lead to corrective action on the basis of any significant elevation of any constituent in any point of compliance well, the impact of an inflated site-wide false positive rate is enormous.

One solution to this problem is to compute a prediction limit that will provide $(1 - \alpha^*)100\%$ confidence of including all k^* future measurements. The simplest approach to

this problem is through use of the Bonferroni inequality [see Miller, 1966; Chew, 1968], noting that from (1.6)

$$\alpha = \frac{\alpha^*}{k^*} . \tag{1.7}$$

Application of (1.7) reveals that in order to have a site-wide error rate at $\alpha^* = .05$ when $k^* = 10$ comparisons are made requires that we test each comparison at the $\alpha = .005\%$ level. The $(1-\alpha)100\%$ prediction limit for the next k^* measurements from a normal distribution is therefore

$$\bar{x} + t_{[n-1,1-\alpha^*/k^*]} s \sqrt{1 + \frac{1}{n}} . \tag{1.8}$$

Table 1.2: One-Sided Values of Student's t Statistic 95% Overall Confidence for Background $n = 4$ to 100 and $k^* = 4$ to 50 Future Measurements

	k^* = Total Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	4.54	5.84	6.74	7.45	8.05	8.57	9.04	9.46	9.85	10.21
8	3.00	3.50	3.81	4.03	4.21	4.35	4.48	4.59	4.69	4.78
12	2.71	3.10	3.32	3.48	3.60	3.71	3.79	3.87	3.93	3.99
16	2.60	2.94	3.14	3.28	3.39	3.48	3.55	3.61	3.67	3.72
20	2.54	2.86	3.04	3.17	3.27	3.35	3.42	3.48	3.53	3.57
24	2.50	2.81	2.98	3.10	3.20	3.27	3.34	3.39	3.44	3.48
28	2.47	2.77	2.94	3.06	3.15	3.22	3.28	3.33	3.38	3.42
32	2.45	2.74	2.91	3.02	3.11	3.18	3.24	3.29	3.33	3.37
36	2.44	2.72	2.88	3.00	3.08	3.15	3.21	3.26	3.30	3.34
40	2.43	2.71	2.87	2.98	3.06	3.13	3.18	3.23	3.27	3.31
44	2.42	2.69	2.85	2.96	3.04	3.11	3.16	3.21	3.25	3.29
48	2.41	2.68	2.84	2.95	3.03	3.09	3.15	3.19	3.24	3.27
52	2.40	2.68	2.83	2.93	3.01	3.08	3.13	3.18	3.22	3.26
56	2.40	2.67	2.82	2.92	3.00	3.07	3.12	3.17	3.21	3.24
60	2.39	2.66	2.81	2.92	3.00	3.06	3.11	3.16	3.20	3.23
64	2.39	2.66	2.81	2.91	2.99	3.05	3.10	3.15	3.19	3.22
68	2.38	2.65	2.80	2.90	2.98	3.04	3.10	3.14	3.18	3.22
72	2.38	2.65	2.80	2.90	2.98	3.04	3.09	3.13	3.17	3.21
76	2.38	2.64	2.79	2.89	2.97	3.03	3.08	3.13	3.17	3.20
80	2.37	2.64	2.79	2.89	2.97	3.03	3.08	3.12	3.16	3.20
84	2.37	2.64	2.78	2.88	2.96	3.02	3.07	3.12	3.16	3.19
88	2.37	2.63	2.78	2.88	2.96	3.02	3.07	3.11	3.15	3.19
92	2.37	2.63	2.78	2.88	2.95	3.01	3.07	3.11	3.15	3.18
96	2.37	2.63	2.77	2.87	2.95	3.01	3.06	3.11	3.14	3.18
100	2.36	2.63	2.77	2.87	2.95	3.01	3.06	3.10	3.14	3.17

Table 1.2 displays one-sided values of $t_{[n-1,\alpha^*/k^*]}$ for $n = 4$ to 100 and $k^* = 5$ to 50 and Table 1.3 displays corresponding two-sided values.

Although the prediction limit in (1.8) limits the probability of any one of k^* future measurements exceeding the limit by chance alone to α^*, it does so at the expense of the false negative rate. To illustrate this point, Figure 1.1, displays statistical power curves for prediction limits for the next k^*= 1, 10, and 50 comparisons based on a background

Table 1.3: Two-Sided Values of Student's t Statistic 95% Overall Confidence for Background $n = 4$ to 100 and $k^* = 4$ to 50 Future Measurements

	k^* = Total Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	5.84	7.45	8.57	9.46	10.21	10.87	11.45	11.98	12.47	12.92
8	3.50	4.03	4.35	4.59	4.78	4.94	5.08	5.20	5.31	5.41
12	3.10	3.48	3.71	3.87	3.99	4.10	4.19	4.26	4.33	4.39
16	2.94	3.28	3.48	3.61	3.72	3.81	3.88	3.95	4.00	4.06
20	2.86	3.17	3.35	3.48	3.57	3.65	3.72	3.78	3.83	3.88
24	2.81	3.10	3.27	3.39	3.48	3.56	3.62	3.67	3.72	3.76
28	2.77	3.06	3.22	3.33	3.42	3.49	3.55	3.60	3.65	3.69
32	2.74	3.02	3.18	3.29	3.37	3.44	3.50	3.55	3.59	3.63
36	2.72	3.00	3.15	3.26	3.34	3.41	3.46	3.51	3.55	3.59
40	2.71	2.98	3.13	3.23	3.31	3.38	3.43	3.48	3.52	3.56
44	2.69	2.96	3.11	3.21	3.29	3.35	3.41	3.45	3.49	3.53
48	2.68	2.95	3.09	3.19	3.27	3.34	3.39	3.43	3.47	3.51
52	2.68	2.93	3.08	3.18	3.26	3.32	3.37	3.42	3.46	3.49
56	2.67	2.92	3.07	3.17	3.24	3.31	3.36	3.40	3.44	3.48
60	2.66	2.92	3.06	3.16	3.23	3.30	3.35	3.39	3.43	3.46
64	2.66	2.91	3.05	3.15	3.22	3.29	3.34	3.38	3.42	3.45
68	2.65	2.90	3.04	3.14	3.22	3.28	3.33	3.37	3.41	3.44
72	2.65	2.90	3.04	3.13	3.21	3.27	3.32	3.36	3.40	3.43
76	2.64	2.89	3.03	3.13	3.20	3.26	3.31	3.35	3.39	3.42
80	2.64	2.89	3.03	3.12	3.20	3.26	3.31	3.35	3.38	3.42
84	2.64	2.88	3.02	3.12	3.19	3.25	3.30	3.34	3.38	3.41
88	2.63	2.88	3.02	3.11	3.19	3.25	3.29	3.34	3.37	3.41
92	2.63	2.88	3.01	3.11	3.18	3.24	3.29	3.33	3.37	3.40
96	2.63	2.87	3.01	3.11	3.18	3.24	3.29	3.33	3.36	3.40
100	2.63	2.87	3.01	3.10	3.17	3.23	3.28	3.32	3.36	3.39

sample of $n = 8$ measurements, setting the individual comparison false positive rate to $\alpha = .05/k^*$. In Figure 1.1, contamination was introduced into a single monitoring well for a single constituent; hence, only one of 1, 10, or 50 comparisons was contaminated. The power curves in Figure 1.1 therefore display the probability of detecting a very localized release that impacts only one of k^* future measurements. In practice, we would expect contamination to impact several wells and constituents, therefore the probability estimates in Figure 1.1 represent a lower bound. Inspection of Figure 1.1 reveals that the false positive rates for $k^* = 1$, 10, and 50 future comparisons all approach the nominal level of 5%; however, false negative rates are dramatically affected by adjusting for larger numbers of future comparisons. For a difference of four standard deviation units and eight background samples, the false negative rates are 4%, 39%, and 57% for $k^* = 1$, 10, and 50, respectively (i.e., 1 minus the probability of a significant result at x-axis = 4 sd units). These results indicate that by simply performing a statistical adjustment to the prediction limit to provide an overall site-wide false positive rate not greater than 5%, we sacrifice the false negative rate (i.e., failure to detect contamination when present), an unacceptable outcome. Control of the false negative rate at the expense of the false positive rate is also unacceptable. For example, the original Resource Conservation and Recovery Act (RCRA) regulation required that quadruplicate samples be obtained (i.e., a single sample split into four aliquots), and replicated measurements were to be treated as if they were independent. Of course, the

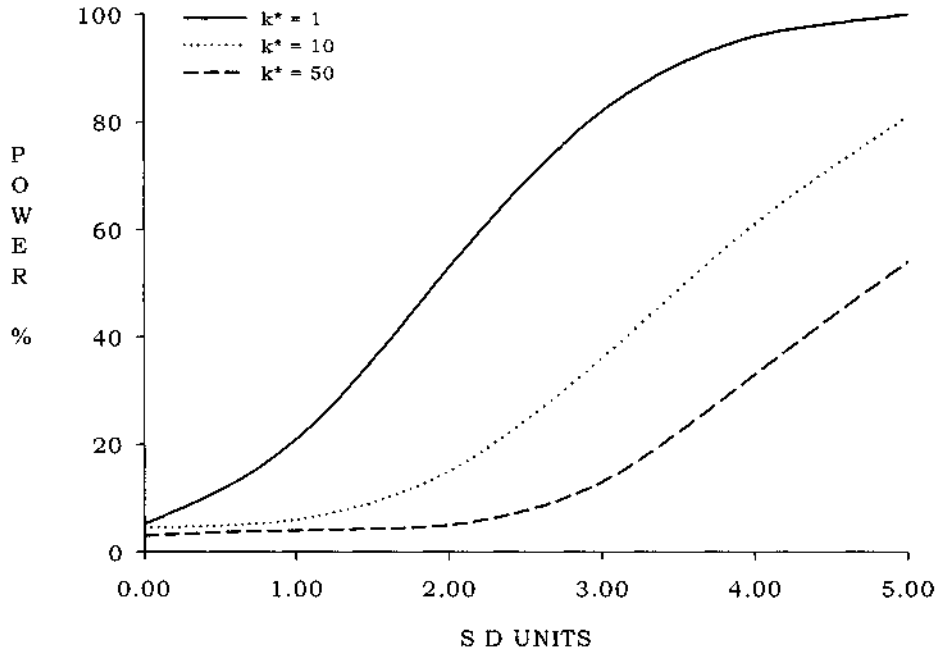

Figure 1.1: Power of 95% Confidence Bonferroni Normal Prediction Limits for 1, 10, and 50 Future Comparisons

intra-sample correlation (i.e., correlation among the measurements for the four aliquots) was typically near unity and not zero, as assumed by the Cochran's Approximation to the Behrens-Fisher (CABF) t-statistic. As such, the false positive rate approached 100%. A test with a false positive rate of 100% has a false negative rate of zero since it will trigger a site assessment regardless of the data.

Example 1.2

Returning to the TOC example dataset in Table 1.1, we may wish to apply the computed prediction limit to new TOC measurements from each of 10 downgradient monitoring wells, limiting the overall probability of *any* false positive result to 5%. To do this, we note that $\alpha^*/k^* = .05/10 = .005$ and the 99.5% upper percentage point of Student's t-distribution with 7 degrees of freedom is $t_{[7,.005]} = 3.499$. The upper prediction limit is therefore

$$11.0 + 3.499(0.61)\sqrt{1 + \frac{1}{8}} = 13.26 \text{ mg/L},$$

in contrast to 12.23 mg/L for a single future comparison.

1.4 NORMAL PREDICTION LIMITS WITH RESAMPLING

The best currently available approach to balancing false positive and false negative rates in groundwater monitoring applications is through the use of verification resampling. Here, in the event of an initial exceedance, one or more verification resamples are obtained and a statistical exceedance is declared if some number of the resampled values also exceeds the limit. In small monitoring programs, it is sometimes possible to declare an exceedance if any of the resampled values exceed the limit [see Gibbons, 1991a]. Alternatively, when background sample sizes are small and the number of future comparisons is large, a reasonable balance between false positive and false negative rates may require that statistical exceedance is declared only if all resampled values exceed the limit [see Davis and McNichols, 1987; Gibbons, 1990]. For this reason, it is critical that the number of monitoring wells and constituents (i.e., k^*) be carefully selected and kept to a minimum.

To illustrate the effects of verification resampling on the false positive rate of a test in which the individual test-wise false positive rate is set at $\alpha = .01$, consider a site with $k^* = 50$ future comparisons and one verification resample. Assuming independence among the k^* future comparisons,

$$\begin{aligned} \alpha^* &= 1 - \Pr(\text{all wells ok}) \\ &= 1 - (\Pr(\text{one well ok}))^{k^*} \\ &= 1 - (1 - \alpha + \alpha(1 - \alpha))^{k^*} \\ &= 1 - (1 - .01 + .01(1 - .01))^{50} \\ &= .005. \end{aligned} \tag{1.9}$$

In this equation, the first $1 - \alpha$ is for the initial sample being in bounds and the $\alpha(1 - \alpha)$ is for the initial sample out of bounds but the resample in bounds. In this case, the verification resample has allowed us to use a 99% confidence prediction limit for 50 future measurements. Without verification resampling, we could have only provided a site-wide 95% confidence level for $k^* = 5$ future monitoring measurements (i.e., $\alpha = \alpha^*/k^* = .05/5 = .01$), using exactly the same individual test-level false positive rate (i.e., $\alpha = .01$) and the corresponding prediction limit.

Now, consider a monitoring program in which, in the event of an initial significant increase, two verification resamples are to be obtained and a significant result is recorded only if both verification resamples exceed the limit. In this case, the site-wide false positive rate is

$$\begin{aligned} \alpha^* &= 1 - (1 - \alpha^3)^{k^*} \\ &= 1 - (1 - .01^3)^{50} \\ &= .00005, \end{aligned} \tag{1.10}$$

which is the probability of failing at least one of the initial samples and both of the verification resamples. This result suggests that for this example, we have gone too far in that the site-wide false positive rate is now well below the nominal 5% level.

As a more conservative alternative, consider a monitoring program, which, in the event of an initial exceedance, two verification resamples are obtained and a significant exceedance is recorded if either resampled value exceeds the limit. In this case, the site-wide false positive rate is given by

$$\begin{aligned} \alpha^* &= 1-(1-\alpha+\alpha(1-\alpha)^2)^{k^*} \\ &= 1-(1-.01+.01(1-.01)^2)^{50} \\ &= .01, \end{aligned} \tag{1.11}$$

i.e., the product of failing an initial sample and at least one of the two verification resamples.

In any of these cases, we should select the most powerful solution that provides a reasonable site-wide false positive rate (i.e., $\alpha^* \sim .05$) within budgetary, legislative, and independence constraints. To do this, select Equations (1.9), (1.10), or (1.11) such that $\alpha^* \sim .05$ for $\alpha \sim .01$ (i.e., site-wide false positive rate of approximately 5% and an individual test false positive rate of approximately 1%). In this way, a reasonable balance between false positive and false negative rates is achieved. Note, however, that these computations require the monitoring samples and resamples to be (adequately) stochastically independent. This implies a certain minimum time between samples. For quarterly monitoring, at most two resamples are reasonable.

Example 1.3

Returning to the TOC example dataset in Table 1.1, we may wish to apply the computed prediction limit to new TOC measurements from each of 10 downgradient monitoring wells for each of five monitoring constituents for a total of 50 future comparisons. Assuming that the five constituents are reasonably independent, the upper 99% confidence normal prediction limit for a single new measurement,

$$11.0+2.998(0.61)\sqrt{1+\frac{1}{8}} = 12.94 \text{ mg/L}.$$

will provide an overall site-wide false positive rate of

$$1-(1-.01)^{50} = .39$$

or 39% without verification resampling,

$$1-(1-.01+.01(1-.01))^{50} = .005$$

or 0.5% with a single verification resample,

$$1-(1-.01+.01(1-.01)^2)^{50} = .01$$

or 1.0% with failure indicated if *either* of two verification resamples fails (i.e., exceeds 12.94),

$$1-(1-.01^3)^{50} = .00005$$

or 0.005% with failure indicated if *both* of two verification resamples fail.

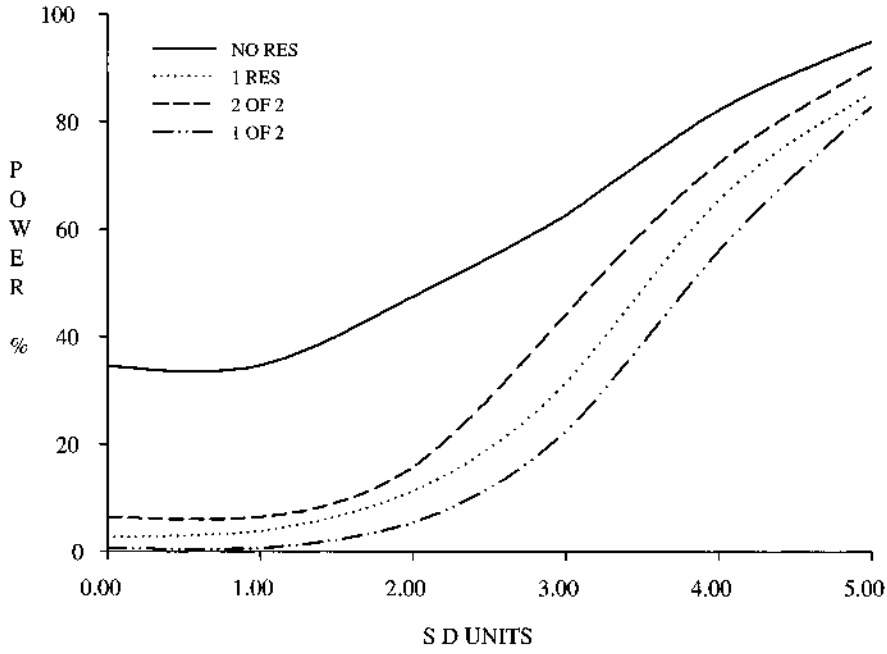

Figure 1.2: Power of 99% Confidence Normal Prediction Limits: 10 Wells and Five Constituents—Four Resampling Plans

To illustrate the effects of resampling on false negative rates, Figure 1.2 displays power curves for the four previously described alternatives (i.e., no resampling, one resample, failing the first and either of two resamples, or failing the first and both of two resamples) using a 99% confidence normal prediction limit.

Figure 1.2 reveals that the plan without resampling has an unacceptably high false positive rate; however, the rate is slightly less than predicted (i.e., 34% versus 39%). The reason for this discrepancy is that the multiple comparisons are not independent, as assumed by the independence-based computations. This problem is discussed in detail in the following section.

Figure 1.2 also reveals that for 10 monitoring wells and 5 constituents and a fixed prediction limit, the best balance between false positive and false negative results is achieved for the plan in which two verification resamples are taken and failure is indicated if either exceeds the limit. Note that the false positive rates obtained via simulation for the various

resampling plans were considerably higher than predicted via the Bonferroni inequality. For example, with a single verification resample we predicted 0.4% false positives but observed 2.7%. For two verification resamples, Plan A produced 6.4% false positives, although only 0.8% were predicted, and Plan B produced 0.6%, although only 0.004% were predicted. These discrepancies are the opposite of what we observed in the case without resampling. The reason for the increased observed false positive rate is that by chance alone, background levels may be particularly low for one well and constituent, but it is this same low background being compared to the verification resample(s). As such, the probability of two successive failures is not the simple product of the individual probabilities, as assumed by the Bonferroni adjustment. Davis and McNichols [1987] have proposed an alternative approach to this problem that overcomes these limitations, which is described in a following section. Caution must be used in comparing false negative rates for tests with different false positive rates. For example, one might conclude from Figure 1.2 that the strategy without resampling has the lowest false negative rate and that this is more important than the fact that it fails by chance alone on one-third of the monitoring events. False negative rates are only meaningful for tests that achieve their intended false positive rates, and comparisons between tests are only appropriate for a fixed false positive rate. The purpose of Figure 1.2 (which clearly violates this advice) is that for a fixed prediction limit (e.g., a 99% confidence normal prediction limit for the next single measurement) applied to different monitoring scenarios (e.g., multiple wells, constituents and resampling strategies), both false positive and false negative rates can vary dramatically. In those cases where the limit is fixed, perhaps by unwise regulation, the verification resampling strategy and the adequacy of the monitoring program as a whole must be based on achieving a balance between false positive and false negative rates. Sacrifice of one for the other is unacceptable.

1.5 SIMULTANEOUS NORMAL PREDICTION LIMITS FOR THE NEXT k SAMPLES

The previous prediction limits for multiple future comparisons are simultaneous in the sense that they control the overall site-wide comparison rate (e.g., $\alpha^* = .05$), assuming that the multiple comparisons are independent. While this is true for a series of intra-well comparisons on a given sampling event (i.e., each well is compared to its own background for a single constituent), it is not true in the context of upgradient versus downgradient comparisons where each new monitoring measurement is compared to the same pooled upgradient background limit. This type of comparison strategy introduces a correlation among the k^* comparisons for each constituent of magnitude $\rho = 1/(n+1)$. An analogous situation occurs in the context of comparing multiple treatment group means to a single control (i.e., Dunnett's test [Dunnett and Sobel, 1955]). The correlation among repeated comparisons makes use of the simple probability product for α^* too conservative. Obtaining the correct joint probability of failure on any one of the k^* comparisons requires evaluation of the equa-correlated multivariate normal distribution, and these probabilities must be integrated over the distribution of s, the sample standard deviation. In this section we define k^* as the number of comparisons (i.e., wells) for a single constituent (i.e., $c = 1$); hence $k^* = k$. We do this so that we can differentiate the effect of multiple comparisons

to a common background from the effects of multiple constituents on the error rates of the overall monitoring program.

Suppose there is interest in comparing k groups with a common background in terms of the means $\bar{X}_0, \bar{X}_1, \ldots, \bar{X}_k$ (and common standard deviation s) of $k+1$ sets of observations which are assumed to be independently and normally distributed; $\bar{X}_0$ refers to the background and $\bar{X}_i$ to the ith comparison group ($i = 1, \ldots, k$) mean. In this case, Dunnett [1955] has provided a procedure for making confidence statements about expected values of k differences $\bar{X}_i - \bar{X}_0$, the procedure having the property that the probability of *all* k statements being simultaneously correct is equal to a specified probability level $1 - \alpha$. Dunnett's procedure and the associated tables were worked out for the case of equal sample sizes in all groups. Here, we will expand the procedure for the case where sample sizes are not equal.

Suppose there are n_0 background measurements, n_1 measurements on the first well, ..., n_k measurements on the kth well, and denote these observations by X_{ij} ($i = 0, 1, \ldots, k$; $j = 1, 2, \ldots, n_i$) and the corresponding ith well mean as $\bar{X}_i$. We assume that the X_{ij} are independent and normally distributed, with common variance σ^2 and means μ_i, and that there is an estimate of σ^2 available (denoted s^2) based on ν degrees of freedom. Let

$$z_i = \frac{\bar{X}_i - \bar{X}_0 - (\mu_i - \mu_0)}{\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}, \tag{1.12}$$

and $t_i = z_i/s$ for $i = 1, 2, \ldots, k$. As Dunnett [1955] notes, the lower confidence limits with joint confidence coefficient $1 - \alpha$ for the k comparison group effects $\mu_i - \mu_0$ are given by

$$\bar{X}_i - \bar{X}_0 - d_i s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}} \tag{1.13}$$

if the k constants d_i are chosen so that

$$\Pr(t_1 < d_1, t_2 < d_2, \ldots, t_k < d_k) = 1 - \alpha. \tag{1.14}$$

To find the k constants d_i that satisfy these equations, joint distribution of the t_i is required, which is the multivariate analogue of Student's t distribution defined by Dunnett and Sobel [1955]. Dunnett [1955] has shown how the problem of tabulating the multivariate t-distribution can be reduced to the problem of tabulating the corresponding multivariate normal (MVN) distribution. For this, note that the joint distribution of the z_i is a MVN distribution with means 0 and variances σ^2. The correlation between z_i and z_j is given by

$$\rho_{ij} = 1 \Big/ \sqrt{\left(\frac{n_0}{n_i} + 1\right)\left(\frac{n_0}{n_j} + 1\right)}. \tag{1.15}$$

Notice that the joint probability statement given above can be written in the following way:

$$\begin{aligned} 1 - \alpha &= \Pr(t_1 < d_1, t_2 < d_2, \ldots, t_k < d_k) \\ &= \Pr(z_1 < d_1 s, z_2 < d_2 s, \ldots, z_k < d_k s) \\ &= \int_{-\infty}^{+\infty} F(d_1 s, d_2 s, \ldots, d_k s)\, f(s)\, ds, \end{aligned} \tag{1.16}$$

where $F(d_1 s, d_2 s, \ldots, d_k s)$ is the MVN cumulative distribution function of the z_i and $f(s)$ is the probability density function of s. Thus, with probability values for $F(\cdot)$, the above equation can be evaluated using numerical integration over the distribution of s. For this, note that the density function of s is given by Pearson and Hartley [1976] as

$$f(s) = \frac{\nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) 2^{(\nu/2)-1}} \sigma^{-\nu} s^{\nu-1} \exp(-\nu s^2/2\sigma^2). \tag{1.17}$$

Since $s^2/\sigma^2 = \chi^2/\nu$, we can rewrite the equation for $1-\alpha$ in terms of integration over the distribution of $y = s/\sigma$ (which is defined on 0 to $+\infty$) as

$$\begin{aligned} 1-\alpha &= \int_0^{+\infty} F(d_1 y, d_2 y, \ldots, d_k y)\, f(y) \frac{\mathrm{d}s}{\mathrm{d}y}\, \mathrm{d}y \\ &= \int_0^{+\infty} F(d_1 y, d_2 y, \ldots, d_k y)\, \frac{\nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) 2^{(\nu/2)-1}} y^{\nu-1} \exp(-\nu y^2/2)\, \mathrm{d}y. \end{aligned} \tag{1.18}$$

Numerical integration over the distribution of y can then be performed to yield the associated probability $1-\alpha$ for selected values of d, k, and ν.

In the present context, we are interested in comparing k new individual measurements (e.g., TOC levels in 10 downgradient monitoring wells) with a collection of n background measurements, perhaps obtained from monitoring wells upgradient of the facility. In this case, $n_i = 1, i = 1, \ldots, k$, $n_0 = n$, and the constant correlation is $\rho_{ij} = 1/(n+1)$. As a result, the need for this correction decreases with increasing background sample size n since as n increases, ρ_{ij} goes to zero. In this special case the probability integral simplifies to

$$F_k(ds, ds, \ldots, ds; \rho) = \int_{-\infty}^{\infty} \left[F^k \left(\frac{ds + \rho^{1/2} y}{\sqrt{1-\rho}} \right) \right] f(y) \mathrm{d}y, \tag{1.19}$$

where $f(\tau) = exp(-\frac{1}{2}\tau^2)/(2\pi)^{1/2}$ and $F(\tau) = \int_{-\infty}^{ds} f(\tau) \mathrm{d}\tau$; see Gupta [1963].

Table 1.4: Dunnett-Type Multivariate t-Statistics 95% Overall Confidence for Background $n = 4$ to 100 and $k^* = 4$ to 50 Future Measurements

	k^* = Total Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	4.00	4.72	5.14	5.42	5.64	5.82	5.96	6.09	6.20	6.29
8	2.90	3.31	3.56	3.72	3.85	3.95	4.04	4.12	4.18	4.24
12	2.67	3.02	3.22	3.36	3.47	3.56	3.63	3.69	3.75	3.80
16	2.57	2.89	3.08	3.21	3.30	3.38	3.45	3.51	3.56	3.60
20	2.51	2.82	3.00	3.12	3.21	3.29	3.35	3.40	3.45	3.49
24	2.48	2.78	2.94	3.06	3.15	3.22	3.28	3.33	3.38	3.42
28	2.45	2.74	2.91	3.02	3.11	3.18	3.24	3.29	3.33	3.37
32	2.44	2.72	2.88	2.99	3.08	3.14	3.20	3.25	3.29	3.33
36	2.42	2.70	2.86	2.97	3.05	3.12	3.18	3.22	3.27	3.30
40	2.41	2.69	2.84	2.95	3.03	3.10	3.15	3.20	3.24	3.28
44	2.40	2.68	2.83	2.94	3.02	3.08	3.14	3.18	3.23	3.26
48	2.39	2.67	2.82	2.93	3.01	3.07	3.12	3.17	3.21	3.25
52	2.39	2.66	2.81	2.92	2.99	3.06	3.11	3.16	3.20	3.23
56	2.38	2.65	2.80	2.91	2.99	3.05	3.10	3.15	3.19	3.22
60	2.38	2.65	2.80	2.90	2.98	3.04	3.09	3.14	3.18	3.21
64	2.38	2.64	2.79	2.89	2.97	3.03	3.09	3.13	3.17	3.20
68	2.37	2.64	2.79	2.89	2.96	3.03	3.08	3.12	3.16	3.20
72	2.37	2.63	2.78	2.88	2.96	3.02	3.07	3.12	3.16	3.19
76	2.37	2.63	2.78	2.88	2.95	3.02	3.07	3.11	3.15	3.19
80	2.36	2.63	2.77	2.87	2.95	3.01	3.06	3.11	3.15	3.18
84	2.36	2.62	2.77	2.87	2.95	3.01	3.06	3.10	3.14	3.18
88	2.36	2.62	2.77	2.87	2.94	3.00	3.05	3.10	3.14	3.17
92	2.36	2.62	2.76	2.86	2.94	3.00	3.05	3.09	3.13	3.17
96	2.36	2.62	2.76	2.86	2.94	3.00	3.05	3.09	3.13	3.16
100	2.35	2.61	2.76	2.86	2.93	2.99	3.05	3.09	3.13	3.16

To aid in application, Table 1.4 provides the constants d for background sample sizes n from 4 to 50 and number of future comparisons k from 2 to 50 for $1 - \alpha = 0.95$. These coefficients may be used in deriving prediction limits of the form

$$\bar{x} + d_{(n,k)} s \sqrt{1 + \frac{1}{n}}. \tag{1.20}$$

Example 1.4

Returning to the TOC example dataset in Table 1.1 and Example 1.2, in which the prediction limit for TOC is to be applied to new TOC measurements from each of 10 downgradient monitoring wells, limiting the overall probability of *any* false positive result to 5%. Using the Bonferroni adjustment, which assumes independence, the value in Table 1.2 that corresponds to $n = 8$ and $k = 10$ is 3.5, leading to the upper prediction limit

$$11.0 + 3.50(0.61)\sqrt{1 + \frac{1}{8}} = 13.26 \text{ mg/L}.$$

In contrast, using the Dunnett-type multivariate t-statistic from Table 1.4 leads to the upper prediction limit

$$11.0 + 3.31(0.61)\sqrt{1 + \frac{1}{8}} = 13.14 \text{ mg/L},$$

which is slightly more conservative. Comparison of Tables 1.2 and 1.4 reveals that the multivariate t and Bonferroni adjusted t statistics become relatively indistinguishable for background sample sizes of $n = 20$ or more, but the difference is more pronounced for a small background sample size. While this discussion helps fix ideas related to multiple comparisons to a common background, it has not been the focus of much if any application in groundwater monitoring programs. This is due largely to the fact that it does not involve verification resampling. A more complete generalization of these foundational ideas, which has enjoyed widespread usage in practical applications in environmental monitoring programs, is described in the following section.

1.6 SIMULTANEOUS NORMAL PREDICTION LIMITS FOR THE NEXT r OF m MEASUREMENTS AT EACH OF k MONITORING WELLS

Davis and McNichols [1987] have generalized the solution given in the previous section to the case in which r out of m samples in each of k future monitoring wells are required in bounds. For example, a detection monitoring program that requires passage of both of two resamples in the event of an initial exceedance is similar to a prediction limit for two of three samples in bounds in each of k wells. Strictly speaking, however, we are concerned with the case in which the first or next two samples are in bounds, which requires a slight modification of their original work. In fact, the multivariate t-statistics in the previous section (i.e., Dunnett's test) represent a special case in which $r = m = 1$.

The derivation is somewhat complicated, but a few key features are described. As in the previous derivations, we assume that the background observations and new monitoring measurements are drawn from the same normal distribution $N(\mu, \sigma^2)$. Expressing $y_{ij} = x_{ij} - \bar{x}$ (i.e., a mean deviation) for $i = 1, \ldots, k$ wells and $j = 1, \ldots, m$ samples and letting $y_{i(r)}$ denote the rth smallest of the y_{ij} for well i, and $y^* = max_i(y_{i(r)})$, then having at least r of m future observations below $\bar{x} + Ks$ is equivalent to $y^* < Ks$, where K is the multiplier we seek. Davis and McNichols [1987] have shown that

$$\begin{aligned} \Pr(y^* < Ks) &= \int_{-\infty}^{\infty} T_{n-1,\sqrt{n}z^*}(\sqrt{n}K) \\ &\quad \times k\left[\int_{-\infty}^{z^*} m\binom{m-1}{r-1}\Phi^{r-1}(t)\phi(t)[1-\Phi(t)^{m-r}]dt\right]^{k-1} \\ &\quad \times m\binom{m-1}{r-1}\Phi^{r-1}(z^*)\phi(z^*)[1-\Phi(z^*)]^{m-r}dz^*, \end{aligned} \tag{1.21}$$

where $T_{\nu,\delta}(\cdot)$ is the cumulative density function of the noncentral t distribution, z^* is the maximum rth order statistic across all k wells, and ϕ and Φ are the standard normal probability density and cumulative distribution functions respectively. The equation is then solved for K such that the right-hand side is equal to $1 - \alpha$. Davis and McNichols [1987] describe a numerical algorithm for obtaining values of K conditional on values of k, r, and m such that the overall confidence level is $1 - \alpha$. The original publication, however, was limited in terms of combinations of k, r, m, and α so routine application of this methodology was generally not possible. Tables 1.5 – 1.13 (prepared by C. Davis)

overcome this limitation in three ways. First, the tables include three levels of α which correspond to Bonferroni adjusted site-wide confidence levels for monitoring programs consisting of 1 ($\alpha = .05$), 10 ($\alpha = .005$) and 20 ($\alpha = .0025$) constituents. Second, the tables provide values of K for background sample sizes of n = 4 to 100 and k = 5 to 50 monitoring wells. Third, separate tables are prepared for one resample and two resamples under Plans A (pass of all resamples) and B (pass at least one resample). In this way, these important results can be practically applied to the widest variety of monitoring programs. Tables 1.5 to 1.7 provide factors of K for the three verification resampling strategies (pass one of two, one of three, and two of three) for $\alpha = .05$, Tables 1.8 to 1.10 for $\alpha = .005$ and Tables 1.11 to 1.13 for $\alpha = .0025$. The importance of disentangling the number of monitoring wells from the number of constituents has often been poorly understood in practice. For inter-well comparisons (i.e., upgradient versus downgradient), the comparisons are correlated, and that correlation has a very specific form, as described in the previous section. While multiple constituents are also likely to be correlated, their intercorrelation is of a much more general form and may not be reliably estimated from the typically small number of observations that are usually available in groundwater monitoring investigations. As such, we adjust for multiple constituents by adjusting α (i.e., $\alpha = \frac{\alpha^*}{c}$), where c is the number of constituents). Note however, that for intra-well comparisons (i.e., each well compared to its own background), the correlation provided by comparing each downgradient well to a pooled upgradient background is no longer present; therefore, all adjustment for multiple comparisons can be achieved by adjusting α. (e.g., $\alpha = \frac{\alpha^*}{kc}$, where k is the number of monitoring wells and c is the number of constituents).

Table 1.5: Simultaneous Normal Prediction Limit Factors for $\alpha = .05$ and One of Two Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	2.47	3.00	3.31	3.52	3.69	3.82	3.94	4.03	4.12	4.19
8	1.72	2.03	2.21	2.33	2.43	2.51	2.57	2.63	2.67	2.72
12	1.56	1.82	1.97	2.07	2.15	2.21	2.27	2.31	2.36	2.39
16	1.48	1.72	1.86	1.95	2.03	2.08	2.13	2.18	2.21	2.25
20	1.44	1.67	1.80	1.89	1.95	2.01	2.06	2.10	2.13	2.16
24	1.41	1.63	1.76	1.84	1.91	1.96	2.01	2.05	2.08	2.11
28	1.39	1.61	1.73	1.81	1.88	1.93	1.97	2.01	2.04	2.07
32	1.38	1.59	1.71	1.79	1.85	1.90	1.95	1.98	2.02	2.04
36	1.37	1.58	1.69	1.77	1.84	1.89	1.93	1.96	1.99	2.02
40	1.36	1.56	1.68	1.76	1.82	1.87	1.91	1.95	1.98	2.00
44	1.35	1.56	1.67	1.75	1.81	1.86	1.90	1.93	1.96	1.99
48	1.34	1.55	1.66	1.74	1.80	1.85	1.89	1.92	1.95	1.98
52	1.34	1.54	1.66	1.73	1.79	1.84	1.88	1.91	1.94	1.97
56	1.33	1.54	1.65	1.73	1.79	1.83	1.87	1.91	1.94	1.96
60	1.33	1.53	1.64	1.72	1.78	1.83	1.87	1.90	1.93	1.95
64	1.33	1.53	1.64	1.72	1.77	1.82	1.86	1.89	1.92	1.95
68	1.32	1.52	1.64	1.71	1.77	1.82	1.85	1.89	1.92	1.94
72	1.32	1.52	1.63	1.71	1.77	1.81	1.85	1.88	1.91	1.94
76	1.32	1.52	1.63	1.70	1.76	1.81	1.85	1.88	1.91	1.93
80	1.32	1.52	1.63	1.70	1.76	1.80	1.84	1.88	1.90	1.93
84	1.32	1.51	1.62	1.70	1.76	1.80	1.84	1.87	1.90	1.93
88	1.31	1.51	1.62	1.70	1.75	1.80	1.84	1.87	1.90	1.92
92	1.31	1.51	1.62	1.69	1.75	1.80	1.83	1.87	1.89	1.92
96	1.31	1.51	1.62	1.69	1.75	1.79	1.83	1.86	1.89	1.92
100	1.31	1.51	1.62	1.69	1.75	1.79	1.83	1.86	1.89	1.91

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.6: Simultaneous Normal Prediction Limit Factors for α = .05 and One of Three Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	1.62	2.02	2.25	2.42	2.55	2.65	2.74	2.82	2.88	2.94
8	1.12	1.37	1.51	1.61	1.69	1.75	1.80	1.84	1.88	1.92
12	1.00	1.21	1.34	1.42	1.49	1.54	1.58	1.62	1.65	1.68
16	.94	1.14	1.26	1.33	1.39	1.44	1.48	1.52	1.55	1.58
20	.91	1.10	1.21	1.28	1.34	1.39	1.43	1.46	1.49	1.51
24	.89	1.08	1.18	1.25	1.31	1.35	1.39	1.42	1.45	1.47
28	.87	1.06	1.16	1.23	1.28	1.33	1.36	1.39	1.42	1.45
32	.86	1.04	1.14	1.21	1.27	1.31	1.34	1.37	1.40	1.42
36	.85	1.03	1.13	1.20	1.25	1.29	1.33	1.36	1.38	1.41
40	.85	1.02	1.12	1.19	1.24	1.28	1.32	1.35	1.37	1.39
44	.84	1.02	1.11	1.18	1.23	1.27	1.31	1.34	1.36	1.38
48	.84	1.01	1.11	1.17	1.22	1.27	1.30	1.33	1.35	1.38
52	.83	1.01	1.10	1.17	1.22	1.26	1.29	1.32	1.35	1.37
56	.83	1.00	1.10	1.16	1.21	1.25	1.29	1.32	1.34	1.36
60	.83	1.00	1.09	1.16	1.21	1.25	1.28	1.31	1.33	1.36
64	.83	1.00	1.09	1.16	1.20	1.24	1.28	1.31	1.33	1.35
68	.82	.99	1.09	1.15	1.20	1.24	1.27	1.30	1.33	1.35
72	.82	.99	1.08	1.15	1.20	1.24	1.27	1.30	1.32	1.34
76	.82	.99	1.08	1.15	1.20	1.23	1.27	1.29	1.32	1.34
80	.82	.99	1.08	1.14	1.19	1.23	1.26	1.29	1.32	1.34
84	.82	.99	1.08	1.14	1.19	1.23	1.26	1.29	1.31	1.34
88	.81	.98	1.08	1.14	1.19	1.23	1.26	1.29	1.31	1.33
92	.81	.98	1.07	1.14	1.19	1.23	1.26	1.28	1.31	1.33
96	.81	.98	1.07	1.14	1.18	1.22	1.26	1.28	1.31	1.33
100	.81	.98	1.07	1.14	1.18	1.22	1.25	1.28	1.31	1.33

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.7: Simultaneous Normal Prediction Limit Factors for $\alpha = .05$ and the First or Next Two Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	2.92	3.47	3.78	3.99	4.16	4.29	4.40	4.50	4.58	4.65
8	2.00	2.31	2.48	2.61	2.70	2.78	2.84	2.89	2.94	2.98
12	1.79	2.05	2.20	2.30	2.38	2.44	2.50	2.54	2.58	2.62
16	1.70	1.94	2.07	2.16	2.24	2.29	2.34	2.38	2.42	2.45
20	1.65	1.87	2.00	2.09	2.15	2.21	2.25	2.29	2.33	2.36
24	1.62	1.83	1.95	2.04	2.10	2.15	2.20	2.23	2.27	2.30
28	1.59	1.80	1.92	2.00	2.06	2.11	2.16	2.19	2.22	2.25
32	1.58	1.78	1.90	1.98	2.04	2.09	2.13	2.16	2.19	2.22
36	1.56	1.76	1.88	1.96	2.02	2.06	2.10	2.14	2.17	2.20
40	1.55	1.75	1.86	1.94	2.00	2.05	2.09	2.12	2.15	2.18
44	1.54	1.74	1.85	1.93	1.99	2.03	2.07	2.10	2.13	2.16
48	1.54	1.73	1.84	1.92	1.97	2.02	2.06	2.09	2.12	2.15
52	1.53	1.72	1.83	1.91	1.96	2.01	2.05	2.08	2.11	2.14
56	1.53	1.72	1.83	1.90	1.96	2.00	2.04	2.07	2.10	2.13
60	1.52	1.71	1.82	1.89	1.95	1.99	2.03	2.06	2.09	2.12
64	1.52	1.71	1.81	1.89	1.94	1.99	2.03	2.06	2.09	2.11
68	1.51	1.70	1.81	1.88	1.94	1.98	2.02	2.05	2.08	2.10
72	1.51	1.70	1.81	1.88	1.93	1.98	2.01	2.05	2.07	2.10
76	1.51	1.70	1.80	1.87	1.93	1.97	2.01	2.04	2.07	2.09
80	1.51	1.69	1.80	1.87	1.92	1.97	2.01	2.04	2.06	2.09
84	1.50	1.69	1.80	1.87	1.92	1.97	2.00	2.03	2.06	2.08
88	1.50	1.69	1.79	1.86	1.92	1.96	2.00	2.03	2.06	2.08
92	1.50	1.69	1.79	1.86	1.92	1.96	2.00	2.03	2.05	2.08
96	1.50	1.68	1.79	1.86	1.91	1.96	1.99	2.02	2.05	2.07
100	1.50	1.68	1.79	1.86	1.91	1.95	1.99	2.02	2.05	2.07

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.8: Simultaneous Normal Prediction Limit Factors for α = .005 and One of Two Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	5.86	6.96	7.60	8.06	8.42	8.71	8.95	9.15	9.34	9.50
8	2.98	3.36	3.59	3.75	3.87	3.97	4.06	4.14	4.20	4.26
12	2.51	2.78	2.94	3.06	3.15	3.22	3.28	3.34	3.38	3.43
16	2.31	2.55	2.69	2.78	2.86	2.92	2.97	3.02	3.06	3.09
20	2.21	2.42	2.55	2.64	2.70	2.76	2.81	2.85	2.88	2.91
24	2.14	2.35	2.46	2.55	2.61	2.66	2.70	2.74	2.77	2.80
28	2.10	2.29	2.40	2.48	2.54	2.59	2.63	2.67	2.70	2.73
32	2.07	2.25	2.36	2.44	2.50	2.54	2.58	2.62	2.65	2.67
36	2.04	2.23	2.33	2.40	2.46	2.50	2.54	2.58	2.60	2.63
40	2.02	2.20	2.30	2.38	2.43	2.48	2.51	2.54	2.57	2.60
44	2.01	2.18	2.28	2.35	2.41	2.45	2.49	2.52	2.55	2.57
48	1.99	2.17	2.27	2.34	2.39	2.43	2.47	2.50	2.53	2.55
52	1.98	2.16	2.25	2.32	2.37	2.42	2.45	2.48	2.51	2.53
56	1.98	2.15	2.24	2.31	2.36	2.40	2.44	2.47	2.49	2.52
60	1.97	2.14	2.23	2.30	2.35	2.39	2.42	2.45	2.48	2.50
64	1.96	2.13	2.22	2.29	2.34	2.38	2.41	2.44	2.47	2.49
68	1.95	2.12	2.21	2.28	2.33	2.37	2.40	2.43	2.46	2.48
72	1.95	2.11	2.21	2.27	2.32	2.36	2.40	2.43	2.45	2.47
76	1.94	2.11	2.20	2.27	2.32	2.36	2.39	2.42	2.44	2.47
80	1.94	2.10	2.20	2.26	2.31	2.35	2.38	2.41	2.44	2.46
84	1.94	2.10	2.19	2.26	2.30	2.34	2.38	2.41	2.43	2.45
88	1.93	2.09	2.19	2.25	2.30	2.34	2.37	2.40	2.42	2.45
92	1.93	2.09	2.18	2.25	2.29	2.33	2.37	2.39	2.42	2.44
96	1.93	2.09	2.18	2.24	2.29	2.33	2.36	2.39	2.41	2.44
100	1.92	2.08	2.18	2.24	2.29	2.33	2.36	2.39	2.41	2.43

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.9: Simultaneous Normal Prediction Limit Factors for α = .005 and One of Three Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	4.02	4.83	5.31	5.66	5.93	6.15	6.34	6.50	6.64	6.76
8	2.10	2.39	2.57	2.70	2.79	2.87	2.94	3.00	3.05	3.10
12	1.76	1.98	2.11	2.20	2.28	2.33	2.38	2.43	2.46	2.50
16	1.62	1.82	1.93	2.00	2.07	2.11	2.16	2.19	2.22	2.25
20	1.55	1.72	1.83	1.90	1.95	2.00	2.03	2.07	2.10	2.12
24	1.50	1.67	1.76	1.83	1.88	1.92	1.96	1.99	2.02	2.04
28	1.47	1.63	1.72	1.78	1.83	1.87	1.91	1.93	1.96	1.98
32	1.44	1.60	1.69	1.75	1.80	1.84	1.87	1.90	1.92	1.94
36	1.42	1.58	1.66	1.72	1.77	1.81	1.84	1.87	1.89	1.91
40	1.41	1.56	1.64	1.70	1.75	1.79	1.82	1.84	1.87	1.89
44	1.40	1.54	1.63	1.69	1.73	1.77	1.80	1.82	1.85	1.87
48	1.39	1.53	1.62	1.67	1.72	1.75	1.78	1.81	1.83	1.85
52	1.38	1.52	1.61	1.66	1.71	1.74	1.77	1.80	1.82	1.84
56	1.37	1.52	1.60	1.65	1.70	1.73	1.76	1.78	1.81	1.83
60	1.37	1.51	1.59	1.64	1.69	1.72	1.75	1.77	1.80	1.82
64	1.36	1.50	1.58	1.64	1.68	1.71	1.74	1.77	1.79	1.81
68	1.36	1.50	1.58	1.63	1.67	1.71	1.74	1.76	1.78	1.80
72	1.35	1.49	1.57	1.63	1.67	1.70	1.73	1.75	1.77	1.79
76	1.35	1.49	1.57	1.62	1.66	1.70	1.72	1.75	1.77	1.79
80	1.35	1.48	1.56	1.62	1.66	1.69	1.72	1.74	1.76	1.78
84	1.34	1.48	1.56	1.61	1.65	1.69	1.71	1.74	1.76	1.78
88	1.34	1.48	1.55	1.61	1.65	1.68	1.71	1.73	1.75	1.77
92	1.34	1.47	1.55	1.60	1.65	1.68	1.71	1.73	1.75	1.77
96	1.34	1.47	1.55	1.60	1.64	1.68	1.70	1.73	1.75	1.77
100	1.33	1.47	1.55	1.60	1.64	1.67	1.70	1.72	1.74	1.76

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.10: Simultaneous Normal Prediction Limit Factors for α = .005 and the First or Next Two Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	6.81	7.95	8.61	9.07	9.42	9.71	9.94	10.15	10.33	10.49
8	3.33	3.72	3.95	4.11	4.24	4.34	4.43	4.50	4.57	4.62
12	2.77	3.05	3.21	3.33	3.42	3.49	3.55	3.60	3.65	3.69
16	2.54	2.78	2.91	3.01	3.09	3.15	3.20	3.24	3.28	3.32
20	2.42	2.63	2.75	2.84	2.91	2.96	3.01	3.05	3.08	3.12
24	2.34	2.54	2.65	2.74	2.80	2.85	2.89	2.93	2.96	2.99
28	2.29	2.48	2.59	2.66	2.72	2.77	2.81	2.85	2.88	2.91
32	2.25	2.43	2.54	2.61	2.67	2.71	2.75	2.79	2.82	2.84
36	2.22	2.40	2.50	2.57	2.63	2.67	2.71	2.74	2.77	2.80
40	2.20	2.37	2.47	2.54	2.60	2.64	2.68	2.71	2.73	2.76
44	2.18	2.35	2.45	2.52	2.57	2.61	2.65	2.68	2.71	2.73
48	2.16	2.33	2.43	2.50	2.55	2.59	2.62	2.66	2.68	2.71
52	2.15	2.32	2.41	2.48	2.53	2.57	2.61	2.64	2.66	2.68
56	2.14	2.31	2.40	2.46	2.51	2.56	2.59	2.62	2.64	2.67
60	2.13	2.29	2.39	2.45	2.50	2.54	2.58	2.60	2.63	2.65
64	2.12	2.29	2.38	2.44	2.49	2.53	2.56	2.59	2.62	2.64
68	2.12	2.28	2.37	2.43	2.48	2.52	2.55	2.58	2.61	2.63
72	2.11	2.27	2.36	2.42	2.47	2.51	2.54	2.57	2.60	2.62
76	2.11	2.26	2.35	2.42	2.46	2.50	2.54	2.56	2.59	2.61
80	2.10	2.26	2.35	2.41	2.46	2.50	2.53	2.56	2.58	2.60
84	2.10	2.25	2.34	2.40	2.45	2.49	2.52	2.55	2.57	2.59
88	2.09	2.25	2.34	2.40	2.44	2.48	2.52	2.54	2.57	2.59
92	2.09	2.24	2.33	2.39	2.44	2.48	2.51	2.54	2.56	2.58
96	2.08	2.24	2.33	2.39	2.44	2.47	2.50	2.53	2.56	2.58
100	2.08	2.24	2.32	2.38	2.43	2.47	2.50	2.53	2.55	2.57

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.11: Simultaneous Normal Prediction Limit Factors for α = .0025 and One of Two Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	7.45	8.82	9.64	10.22	10.66	11.03	11.33	11.59	11.82	12.02
8	3.40	3.82	4.06	4.24	4.37	4.49	4.58	4.66	4.73	4.80
12	2.80	3.08	3.25	3.37	3.46	3.54	3.60	3.66	3.71	3.75
16	2.56	2.79	2.93	3.03	3.11	3.17	3.23	3.27	3.31	3.35
20	2.43	2.64	2.77	2.86	2.92	2.98	3.03	3.07	3.10	3.14
24	2.35	2.55	2.66	2.75	2.81	2.86	2.90	2.94	2.97	3.00
28	2.30	2.49	2.60	2.67	2.73	2.78	2.82	2.86	2.89	2.91
32	2.26	2.44	2.54	2.62	2.68	2.72	2.76	2.79	2.82	2.85
36	2.23	2.40	2.51	2.58	2.63	2.68	2.72	2.75	2.78	2.80
40	2.20	2.38	2.48	2.55	2.60	2.64	2.68	2.71	2.74	2.76
44	2.19	2.36	2.45	2.52	2.57	2.62	2.65	2.68	2.71	2.73
48	2.17	2.34	2.43	2.50	2.55	2.59	2.63	2.66	2.69	2.71
52	2.16	2.32	2.42	2.48	2.53	2.57	2.61	2.64	2.67	2.69
56	2.15	2.31	2.40	2.47	2.52	2.56	2.59	2.62	2.65	2.67
60	2.14	2.30	2.39	2.46	2.50	2.55	2.58	2.61	2.63	2.66
64	2.13	2.29	2.38	2.44	2.49	2.53	2.57	2.60	2.62	2.64
68	2.12	2.28	2.37	2.43	2.48	2.52	2.56	2.58	2.61	2.63
72	2.11	2.27	2.36	2.43	2.47	2.51	2.55	2.57	2.60	2.62
76	2.11	2.27	2.36	2.42	2.47	2.51	2.54	2.57	2.59	2.61
80	2.10	2.26	2.35	2.41	2.46	2.50	2.53	2.56	2.58	2.60
84	2.10	2.26	2.34	2.41	2.45	2.49	2.52	2.55	2.58	2.60
88	2.10	2.25	2.34	2.40	2.45	2.49	2.52	2.54	2.57	2.59
92	2.09	2.25	2.33	2.40	2.44	2.48	2.51	2.54	2.56	2.58
96	2.09	2.24	2.33	2.39	2.44	2.47	2.51	2.53	2.56	2.58
100	2.09	2.24	2.33	2.39	2.43	2.47	2.50	2.53	2.55	2.57

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.12: Simultaneous Normal Prediction Limit Factors for $\alpha = .0025$ and One of Three Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	5.14	6.14	6.75	7.19	7.53	7.80	8.04	8.24	8.42	8.58
8	2.42	2.74	2.93	3.07	3.17	3.26	3.33	3.40	3.45	3.50
12	1.99	2.22	2.35	2.45	2.52	2.58	2.63	2.68	2.72	2.75
16	1.82	2.01	2.12	2.20	2.26	2.31	2.36	2.39	2.43	2.46
20	1.73	1.90	2.00	2.07	2.13	2.17	2.21	2.24	2.27	2.30
24	1.67	1.83	1.93	1.99	2.04	2.09	2.12	2.15	2.18	2.20
28	1.63	1.78	1.87	1.94	1.99	2.03	2.06	2.09	2.11	2.14
32	1.60	1.75	1.84	1.90	1.94	1.98	2.01	2.04	2.07	2.09
36	1.58	1.72	1.81	1.87	1.91	1.95	1.98	2.01	2.03	2.05
40	1.56	1.70	1.79	1.84	1.89	1.92	1.95	1.98	2.00	2.02
44	1.55	1.69	1.77	1.83	1.87	1.90	1.93	1.96	1.98	2.00
48	1.53	1.67	1.75	1.81	1.85	1.89	1.92	1.94	1.96	1.98
52	1.52	1.66	1.74	1.80	1.84	1.87	1.90	1.93	1.95	1.97
56	1.52	1.65	1.73	1.79	1.83	1.86	1.89	1.91	1.93	1.95
60	1.51	1.64	1.72	1.78	1.82	1.85	1.88	1.90	1.92	1.94
64	1.50	1.64	1.71	1.77	1.81	1.84	1.87	1.89	1.91	1.93
68	1.50	1.63	1.71	1.76	1.80	1.83	1.86	1.88	1.91	1.92
72	1.49	1.63	1.70	1.75	1.79	1.83	1.85	1.88	1.90	1.92
76	1.49	1.62	1.70	1.75	1.79	1.82	1.85	1.87	1.89	1.91
80	1.48	1.62	1.69	1.74	1.78	1.81	1.84	1.87	1.89	1.90
84	1.48	1.61	1.69	1.74	1.78	1.81	1.84	1.86	1.88	1.90
88	1.48	1.61	1.68	1.73	1.77	1.81	1.83	1.86	1.88	1.89
92	1.48	1.61	1.68	1.73	1.77	1.80	1.83	1.85	1.87	1.89
96	1.47	1.60	1.68	1.73	1.77	1.80	1.82	1.85	1.87	1.88
100	1.47	1.60	1.67	1.72	1.76	1.79	1.82	1.84	1.86	1.88

Prepared by Charles Davis based on results in Davis and McNichols [1987]

Table 1.13: Simultaneous Normal Prediction Limit Factors for $\alpha = .0025$ and the First or Next Two Samples Inbounds (Factors K Where the Prediction Limit Is $\bar{x} + Ks$)

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	8.65	10.08	10.91	11.48	11.93	12.29	12.59	12.84	13.07	13.27
8	3.78	4.21	4.46	4.64	4.78	4.89	4.98	5.06	5.13	5.20
12	3.07	3.36	3.53	3.65	3.74	3.82	3.88	3.94	3.99	4.03
16	2.78	3.03	3.17	3.26	3.34	3.40	3.46	3.50	3.54	3.58
20	2.64	2.85	2.97	3.06	3.13	3.19	3.23	3.27	3.31	3.34
24	2.54	2.74	2.86	2.94	3.00	3.05	3.09	3.13	3.16	3.19
28	2.48	2.67	2.78	2.85	2.91	2.96	3.00	3.03	3.07	3.09
32	2.44	2.62	2.72	2.79	2.85	2.89	2.93	2.97	2.99	3.02
36	2.40	2.58	2.68	2.75	2.80	2.84	2.88	2.91	2.94	2.97
40	2.37	2.54	2.64	2.71	2.76	2.80	2.84	2.87	2.90	2.92
44	2.35	2.52	2.61	2.68	2.73	2.77	2.81	2.84	2.87	2.89
48	2.33	2.50	2.59	2.66	2.71	2.75	2.78	2.81	2.84	2.86
52	2.32	2.48	2.57	2.64	2.69	2.73	2.76	2.79	2.82	2.84
56	2.31	2.47	2.56	2.62	2.67	2.71	2.74	2.77	2.80	2.82
60	2.30	2.45	2.54	2.61	2.65	2.69	2.73	2.75	2.78	2.80
64	2.29	2.44	2.53	2.59	2.64	2.68	2.71	2.74	2.77	2.79
68	2.28	2.43	2.52	2.58	2.63	2.67	2.70	2.73	2.75	2.77
72	2.27	2.42	2.51	2.57	2.62	2.66	2.69	2.72	2.74	2.76
76	2.26	2.42	2.50	2.56	2.61	2.65	2.68	2.71	2.73	2.75
80	2.26	2.41	2.50	2.56	2.60	2.64	2.67	2.70	2.72	2.74
84	2.25	2.40	2.49	2.55	2.60	2.63	2.66	2.69	2.71	2.74
88	2.25	2.40	2.48	2.54	2.59	2.63	2.66	2.68	2.71	2.73
92	2.24	2.39	2.48	2.54	2.58	2.62	2.65	2.68	2.70	2.72
96	2.24	2.39	2.47	2.53	2.58	2.61	2.64	2.67	2.69	2.71
100	2.24	2.38	2.47	2.53	2.57	2.61	2.64	2.67	2.69	2.71

Prepared by Charles Davis based on results in Davis and McNichols [1987]

To illustrate the strength of this approach, consider a monitoring program with a single monitoring constituent, $k = 50$ monitoring wells and $n = 8$ background measurements. For a single verification resample, the limit is given by $\bar{x} + 2.72s$ (see Table 1.5). For two verification resamples in which an exceedance is recorded if either is exceeded (i.e., Plan A), the limit is given by $\bar{x} + 2.98s$ (see Table 1.7). For two verification resamples in which an exceedance is recorded only if both are exceeded (i.e., Plan B), the limit is given by $\bar{x} + 1.92s$ (see Table 1.6). Note that by requiring both resamples to be exceeded, we can use a smaller value of K and hence have greater power and a lower false negative rate. This result is evident in the power curves displayed in Figure 1.3 This result is contrary to guidance and previous regulations which suggest that passing multiple resamples in the presence of an initial exceedance leads to a more conservative monitoring program. When the correct statistical model is used, in fact the reverse is true.

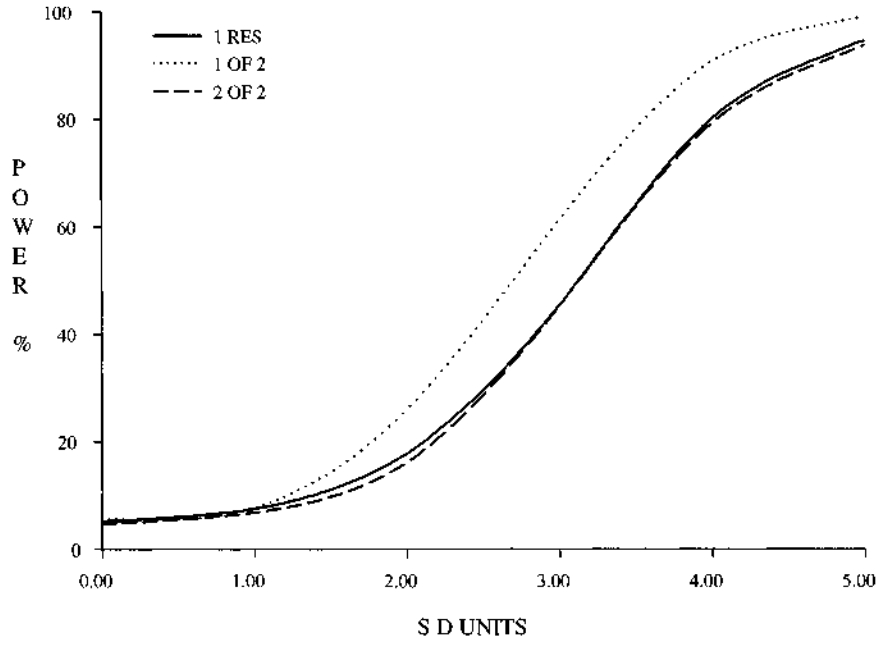

Figure 1.3: Power of 95% Simultaneous Prediction Limits: 10 Wells and Five Constituents—Three Resampling Plans

Finally, also note that despite the difference in multipliers for the three verification resampling plans, the site-wide false positive rate continues to be 5% as intended. For these reasons, simultaneous prediction limits of the form derived by Davis and McNichols [1987] are the optimal choice for normally distributed measurements or those that can be suitably transformed. Subsequent work in this area has extended these limits to other distributions and nonparametric alternatives that do not assume any particular distributional form, both of which will be described in following chapters.

1.7 NORMAL PREDICTION LIMITS FOR THE MEAN(S) OF $m > 1$ FUTURE MEASUREMENTS AT EACH OF k MONITORING WELLS

In certain cases, we may be interested in comparing an average concentration to background. The average concentration may be obtained as (1) the mean of a series of different locations at a given time point, (2) the mean of a series of measurements collected over time at a single location, or (3) the mean of measurements collected over time at a number of locations. While there are a number of statistical approaches for the comparison of mean values for normally distributed random variables (e.g., analysis of variance [ANOVA] or t-tests), these methods pool variance estimates from both background and on-site/downgradient measurements. Since contamination can impact both the absolute magnitude and the variability of the concentration distribution, using potentially impacted data to represent natural variability is not a judicious choice. If there is interest in comparing average on-site/downgradient concentrations to background, then only the background data should be used in deriving the estimate of natural variability. Fortunately, this is exactly the case for prediction limits (i.e., β-expectation tolerance limits; Guttman [1970]) for future mean values. Since the variance of a single measurement σ^2 is much larger than the variance of the mean of m such measurements σ^2/m, the corresponding $(1-\alpha)100\%$ prediction limit for the mean of m future values from a normal distribution,

$$\bar{x} + t_{[n-1,1-\alpha]} s \sqrt{1/m + 1/n}, \tag{1.22}$$

is considerably smaller than the corresponding prediction limit for m future individual measurements. Also, note that since the m measurements are now pooled into a single comparison for their mean, the total number of comparisons is reduced to a single comparison, further reducing the size of the statistical limit estimate. Of course, if we are interested in comparing the mean of m measurements at each of k locations, the Bonferroni adjusted prediction limit becomes

$$\bar{x} + t_{[n-1,1-\alpha/k]} s \sqrt{1/m + 1/n}\,. \tag{1.23}$$

Finally, if we are interested in comparing the grand mean averaging over both time and locations, the normal prediction limit becomes

$$\bar{x} + t_{[n-1,1-\alpha]} s \sqrt{1/km + 1/n}\,, \tag{1.24}$$

which once again reduces to a single comparison for the mean of the km measurements.

While the first and last cases involve a single comparison (per constituent), the second case (i.e., the mean of m samples at each of k locations) produces a nontrivial multiple comparison case in that the k locations are all compared to a common background. In this case, the correlation between the k comparisons is constant (assuming an equal number of measurements per location) with value

$$\rho = 1/\sqrt{\left(\frac{n}{m}+1\right)\left(\frac{n}{m}+1\right)}. \tag{1.25}$$

For example, with $n = 8$ background measurements and $m = 4$ on-site measurements in each of the k locations, the correlation between the comparisons is $\rho = .33$ as compared to $\rho = .11$ for a single ($m = 1$) measurement per well. As such, the coefficients obtained under the independence assumption will be more severely biased (i.e., too large) when comparing

on-site/downgradient means as opposed to individual measurements to background. The variation on Dunnett's test described in the previous section can be applied here as well. The tabled values are different because the correlations change as a function of the number of measurements at each location. Table 1.14 presents the appropriate coefficients for $m = 4$ measurements at each of $k = 4 - 50$ locations. Table 1.15 presents the appropriate coefficients for $m = 8$ measurements at each of $k = 4$ to 50 locations. Table 1.16 presents the appropriate coefficients for $m = 12$ measurements at each of $k = 4$ to 50 locations.

Table 1.14: Dunnett-Type Multivariate t-Statistics $[d(n, m, k)]$ for Comparing the Mean of $m = 4$ Measurements at Each $k = 4$ to 50 Locations Based on $n = 4$ to 100 Background Measurements with 95% Confidence

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	3.70	4.25	4.56	4.78	4.94	5.07	5.18	5.27	5.35	5.42
8	2.81	3.18	3.39	3.54	3.65	3.73	3.81	3.87	3.93	3.97
12	2.62	2.95	3.14	3.27	3.37	3.44	3.51	3.57	3.62	3.66
16	2.54	2.85	3.03	3.15	3.24	3.31	3.38	3.43	3.47	3.52
20	2.49	2.79	2.96	3.08	3.16	3.24	3.30	3.35	3.39	3.43
24	2.46	2.75	2.92	3.03	3.12	3.19	3.24	3.29	3.34	3.38
28	2.44	2.73	2.89	3.00	3.08	3.15	3.21	3.25	3.30	3.34
32	2.43	2.71	2.86	2.97	3.06	3.12	3.18	3.23	3.27	3.30
36	2.41	2.69	2.85	2.95	3.04	3.10	3.16	3.20	3.24	3.28
40	2.40	2.68	2.83	2.94	3.02	3.08	3.14	3.18	3.23	3.26
44	2.40	2.67	2.82	2.93	3.01	3.07	3.12	3.17	3.21	3.25
48	2.39	2.66	2.81	2.92	2.99	3.06	3.11	3.16	3.20	3.23
52	2.38	2.65	2.80	2.91	2.99	3.05	3.10	3.15	3.19	3.22
56	2.38	2.65	2.80	2.90	2.98	3.04	3.09	3.14	3.18	3.21
60	2.38	2.64	2.79	2.89	2.97	3.03	3.08	3.13	3.17	3.20
64	2.37	2.64	2.78	2.89	2.96	3.03	3.08	3.12	3.16	3.20
68	2.37	2.63	2.78	2.88	2.96	3.02	3.07	3.12	3.16	3.19
72	2.37	2.63	2.78	2.88	2.95	3.02	3.07	3.11	3.15	3.18
76	2.36	2.63	2.77	2.87	2.95	3.01	3.06	3.11	3.14	3.18
80	2.36	2.62	2.77	2.87	2.95	3.01	3.06	3.10	3.14	3.17
84	2.36	2.62	2.77	2.87	2.94	3.00	3.05	3.10	3.14	3.17
88	2.36	2.62	2.76	2.86	2.94	3.00	3.05	3.09	3.13	3.17
92	2.36	2.62	2.76	2.86	2.94	3.00	3.05	3.09	3.13	3.16
96	2.35	2.61	2.76	2.86	2.93	2.99	3.04	3.09	3.13	3.16
100	2.35	2.61	2.76	2.86	2.93	2.99	3.04	3.08	3.12	3.16

Table 1.15: Dunnett-Type Multivariate t-Statistics $[d(n, m, k)]$ for Comparing the Mean of $m = 8$ Measurements at Each of $k = 4$ to 50 Locations Based on $n = 4$ to 100 Background Measurements with 95% Confidence

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	3.48	3.92	4.16	4.33	4.46	4.56	4.65	4.72	4.78	4.84
8	2.73	3.05	3.23	3.36	3.45	3.53	3.59	3.64	3.69	3.73
12	2.57	2.88	3.04	3.16	3.25	3.32	3.38	3.43	3.47	3.51
16	2.51	2.80	2.96	3.07	3.16	3.23	3.28	3.33	3.37	3.41
20	2.47	2.75	2.91	3.02	3.10	3.17	3.23	3.28	3.32	3.35
24	2.44	2.72	2.88	2.99	3.07	3.14	3.19	3.24	3.28	3.32
28	2.43	2.70	2.86	2.96	3.04	3.11	3.16	3.21	3.25	3.29
32	2.41	2.69	2.84	2.95	3.03	3.09	3.14	3.19	3.23	3.27
36	2.40	2.67	2.83	2.93	3.01	3.07	3.13	3.17	3.21	3.25
40	2.39	2.66	2.81	2.92	3.00	3.06	3.11	3.16	3.20	3.23
44	2.39	2.66	2.81	2.91	2.99	3.05	3.10	3.15	3.19	3.22
48	2.38	2.65	2.80	2.90	2.98	3.04	3.09	3.14	3.18	3.21
52	2.38	2.64	2.79	2.89	2.97	3.03	3.08	3.13	3.17	3.20
56	2.37	2.64	2.79	2.89	2.96	3.03	3.08	3.12	3.16	3.20
60	2.37	2.63	2.78	2.88	2.96	3.02	3.07	3.12	3.15	3.19
64	2.37	2.63	2.78	2.88	2.95	3.01	3.07	3.11	3.15	3.18
68	2.36	2.63	2.77	2.87	2.95	3.01	3.06	3.11	3.14	3.18
72	2.36	2.62	2.77	2.87	2.94	3.01	3.06	3.10	3.14	3.17
76	2.36	2.62	2.77	2.87	2.94	3.00	3.05	3.10	3.13	3.17
80	2.36	2.62	2.76	2.86	2.94	3.00	3.05	3.09	3.13	3.16
84	2.36	2.62	2.76	2.86	2.93	3.00	3.05	3.09	3.13	3.16
88	2.35	2.61	2.76	2.86	2.93	2.99	3.04	3.09	3.12	3.16
92	2.35	2.61	2.76	2.85	2.93	2.99	3.04	3.08	3.12	3.15
96	2.35	2.61	2.75	2.85	2.93	2.99	3.04	3.08	3.12	3.15
100	2.35	2.61	2.75	2.85	2.92	2.98	3.04	3.08	3.12	3.15

Table 1.16: Dunnett-Type Multivariate t-Statistics $[d(n, m, k)]$ for Comparing the Mean of $m = 12$ Measurements at Each of $k = 4$ to 50 Locations Based on $n = 4$ to 100 Background Measurements with 95% Confidence

	k = Number of Future Comparisons									
n	5	10	15	20	25	30	35	40	45	50
4	3.34	3.72	3.93	4.07	4.17	4.26	4.33	4.40	4.45	4.50
8	2.66	2.95	3.11	3.22	3.31	3.37	3.43	3.48	3.52	3.56
12	2.53	2.81	2.96	3.07	3.15	3.21	3.27	3.31	3.35	3.39
16	2.47	2.75	2.90	3.01	3.08	3.15	3.20	3.25	3.29	3.32
20	2.44	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.25	3.28
24	2.42	2.69	2.84	2.95	3.02	3.09	3.14	3.18	3.22	3.26
28	2.41	2.68	2.83	2.93	3.01	3.07	3.12	3.17	3.21	3.24
32	2.40	2.67	2.81	2.92	2.99	3.06	3.11	3.15	3.19	3.23
36	2.39	2.66	2.80	2.91	2.98	3.05	3.10	3.14	3.18	3.21
40	2.38	2.65	2.80	2.90	2.97	3.04	3.09	3.13	3.17	3.20
44	2.38	2.64	2.79	2.89	2.97	3.03	3.08	3.12	3.16	3.20
48	2.37	2.64	2.78	2.88	2.96	3.02	3.07	3.12	3.16	3.19
52	2.37	2.63	2.78	2.88	2.95	3.02	3.07	3.11	3.15	3.18
56	2.37	2.63	2.77	2.87	2.95	3.01	3.06	3.11	3.14	3.18
60	2.36	2.62	2.77	2.87	2.95	3.01	3.06	3.10	3.14	3.17
64	2.36	2.62	2.77	2.87	2.94	3.00	3.05	3.10	3.13	3.17
68	2.36	2.62	2.76	2.86	2.94	3.00	3.05	3.09	3.13	3.16
72	2.36	2.62	2.76	2.86	2.93	3.00	3.05	3.09	3.13	3.16
76	2.35	2.61	2.76	2.86	2.93	2.99	3.04	3.09	3.12	3.16
80	2.35	2.61	2.76	2.85	2.93	2.99	3.04	3.08	3.12	3.15
84	2.35	2.61	2.75	2.85	2.93	2.99	3.04	3.08	3.12	3.15
88	2.35	2.61	2.75	2.85	2.92	2.98	3.03	3.08	3.12	3.15
92	2.35	2.61	2.75	2.85	2.92	2.98	3.03	3.08	3.11	3.15
96	2.35	2.60	2.75	2.85	2.92	2.98	3.03	3.07	3.11	3.14
100	2.35	2.60	2.75	2.84	2.92	2.98	3.03	3.07	3.11	3.14

Example 1.5

As an illustration, consider the following hexavalent chromium data and assume, for the purpose of illustration, that they represent background conditions.

Date	Result
08/02/96	29
09/09/96	14
09/30/96	13
10/30/96	14
12/04/96	19
01/06/97	9
02/10/97	<1
05/08/97	33
08/05/97	150
11/11/97	60
02/02/98	57

For the mean of four future samples, the 95% normal upper prediction limit (UPL) is

$$\begin{aligned} UPL &= \bar{x} + ts\sqrt{1/m + 1/n} \\ &= 36.136 + 1.812(42.515)\sqrt{1/4 + 1/11} \\ &= 81.116 \text{ mg/L}. \end{aligned}$$

Since there is a single future comparison (i.e., $k = 1$), the limit is based on Student's t-distribution. Now consider $k = 5$ locations. From Table 1.14 we obtain a multivariate t value of 2.67 and a corresponding UPL of 102.415 mg/L.

1.8 SUMMARY

A variety of statistical prediction limits and intervals have been presented for normally distributed measurements in order of increasing statistical sophistication required for increasingly complex applications. The final prediction limits presented, based on the work of Davis and McNichols [1987], are most appealing since they consider both multiple comparisons and verification resampling in the most statistically rigorous way. Furthermore, they allow us to partition out the effects of multiple monitoring wells from the effects of multiple constituents for inter-well comparisons. The reader should note, however, that in the case of $k > 1$ we are invariably describing an upgradient versus downgradient monitoring plan. For intra-well comparisons we select $k = 1$ and $\alpha = \frac{\alpha^*}{k^*}$, where k^* is the total number of comparisons, in this case the product of wells and constituents. The assumption here is that either no spatial variability exists or that spatial variability in *all* downgradient wells is adequately described and can be statistically modeled by the spatial

variability in the small number of upgradient wells available at the facility (a topic to be discussed in a following chapter). Interestingly, there are no current regulatory requirements on the number of upgradient wells, and to minimize expenses, owners/operators typically keep the number of upgradient wells to fewer than three or four. Often, there is only a single upgradient well, and potential contamination is completely confounded with spatial variability, making the upgradient versus downgradient comparison strategy meaningless. Even with three or four upgradient wells, it is extremely unlikely that spatial variability across the site as a whole will be adequately characterized. The methods described here, and indeed any other approach, will not work in the presence of such spatial variability. If predisposal data exist, then estimates of site-wide spatial variability are available and can be incorporated into prediction limits using components of variance models to be described. Unfortunately, predisposal data are rarely available at most sites despite the fact that their benefits far outweigh their cost. In these cases, intra-well comparisons may be the only viable alternative. Intra-well comparisons are always more powerful if they are justified.