DQO Statistics Bulletin

Statistical Methods for the Data Quality Objectives Process

Volume 1, Number 1, PNL-SA-26377 June 1995

Preface

This issue is the first in a series of bulletins describing statistical methods with demonstrated potential for reducing sampling costs and increasing information content of environmental data and statistical analyses for site assessments, remedial actions, and cleanups at U.S. Department of Energy (DOE) sites. The series is being sent to DOE HQ managers, DOE national laboratory personnel, statisticians in government, academia, and industry, and other technical staff engaged or interested in the design and implementation of cleanup activities at DOE sites across the United States.

The purpose of this effort is two-fold. First, we hope to encourage development of a common knowledge base of efficient statistical methods in order to facilitate the communication, evaluation, and selection of site-specific optimal sampling plans using the U.S. Environmental Protection Agency (EPA) Data Quality Objectives (DQO) process (EPA 1993, 1994). Indeed, promoting and improving the DQO process is the primary motivation for developing this *DQO Statistics Bulletin* series. Second, we want to stimulate the further development and application of statistical methods for environmental remedial actions and cleanups.

Plans are to issue a bulletin every 3 months. Each short, reader-friendly bulletin will include an overview discussion and one or more examples illustrating the featured topic, method, or subject area. Key references will be provided, but no attempt will be made to provide in-depth discussions of underlying statistical theory.

This effort is being implemented by the Pacific Northwest Laboratory (PNL) for DOE's Laboratory Management Division (EM-26). We invite all readers from other DOE national laboratories, government agencies, academia, and industry to participate. If you are interested in collaborating on an issue, please contact any of us at addresses listed on page 2.

--*Richard O. Gilbert*, bulletin manager

Featured in this issue: *Ranked Set Sampling*

Ranked set sampling (RSS) is a potentially very useful statistical sampling design for environmental pollution problems. The method

- *Uses auxiliary information such as judgment, historical data, or new field screening data to limit the amount of more expensive and defendable additional sample data that must be collected to meet established DQOs.*
- *Should be useful if the auxiliary information is already available or can be obtained at very little cost.*

1.0 Introduction

McIntyre (1952) introduced ranked set sampling when he estimated the mean of pasture hay production by using visual estimations of hay production ("prior information") in combination with harvesting and weighing a limited number of plots. The early published papers on RSS also focused on estimating the mean. It has been shown (e.g., Patil et al. 1992b, 1994a, 1994b) that the precision of the mean estimated using RSS is expected to be greater than that of a mean estimated using the same number of measurements obtained using simple random sampling. That is, we get more "bang for the buck." From a cost perspective, RSS has the potential for estimating the mean with required precision and confidence (as specified in the DQOs) at less cost than if data are obtained using simple random sampling.

The *DQO Statistics Bulletin* is produced for the U.S. Department of Energy, Laboratory Management Division, EM-26, as a deliverable of the Data Quality Objectives Training and Support Project. The Pacific Northwest Laboratory (PNL) manages the project for DOE.

We encourage readers to contact us with ideas, questions, comments, and article contributions.

Brent Pulsipher, PNL project manager
Phone 509-375-3989, Fax 509-375-3614
e-mail ba_pulsipher@pnl.gov

Richard O. Gilbert, bulletin manager
Phone 509-375-2979, fax 509-375-3614
e-mail ro_gilbert@pnl.gov

David Bottrell, DOE-HQ program manager
Phone 301-903-7251, fax 301-903-7613
e-mail david.bottrell@mail.er.doe.gov

PNL, a multiprogram national laboratory, is operated for DOE under Contract DE-AC06-76RLO 1830 by Battelle Memorial Institute.

The RSS methodology can be easily applied to environmental pollution data. RSS consists of first selecting m sampling units (e.g., soil samples, 1-m^2 parcels of ground, the air drawn through an air sampler during a 5-minute time period) from the target population (geographical area or time period of interest) using simple random sampling. After these units have been selected, inexpensive auxiliary information is used to rank (order from smallest to largest) the selected units with regard to the expected concentration of the pollutant in each unit. Once the units are ranked, a specific procedure (detailed in Section 2) is used to select a subset of the ranked units. These selected units are then measured for the pollutant of interest. The measurements are then used to estimate the true mean, μ, for the site.

The ranking of sampling units might be accomplished using expert judgment based on knowledge of operational history at the site, previous data obtained at the site, an inexpensive auxiliary measurement, visual inspection of sampling units, or some combination of these methods. Field screening techniques such as portable *in situ* detector or fiber-optic readings may be particularly effective at ranking units if the reading of a unit is a reasonably good indicator of the pollutant concentration in the unit. Patil et al. (1994b) suggest that, at hazardous waste sites, sampling units can be ranked according to their approximate contamination levels as revealed by visual cues such as defoliation and soil discoloration, special chemically responsive papers, electromagnetic readings, and remotely sensed data.

It should be noted that the estimated mean is *not* computed using auxiliary measurements that may be used to do the ranking. Instead, the estimated mean is computed using the usual ("accurate") measurements needed to achieve the quality requirements specified in the DQOs.

Ideally, one would like the ranking process to be perfect, in the sense that the sampling unit given the rank of 1 does indeed have the smallest true concentration of the pollutant among the m units, the unit given the rank of 2 has the next largest true concentration among the m units, ..., and the unit given the rank of m has the largest true concentration among the m units. Fortunately, the ranking process does not need to be perfect for RSS to yield a more precise estimate of the mean than can be obtained using simple random sampling. Patil et al. (1994b) indicate that the relative efficiency (precision) of RSS compared to simple random sampling can be several hundred percent depending upon factors such as ranking accuracy, number of samples, and the underlying population distribution.

In Sections 2 and 3, we describe the RSS method in detail and how it might be used to estimate the mean plutonium concentration in soil at a DOE site. Next, in Section 4, we discuss how to replicate the RSS process. Technical issues and possible application of RSS in environmental studies are covered in Section 5. Future activities with regard to RSS are suggested in Section 6. References cited in this issue are listed in Section 7.

2.0 RSS Method

The RSS method is best illustrated via an example based on an actual remediation site.

At several weapons testing areas on or adjacent to the Nevada Test Site (NTS), the soil within fenced areas is contaminated with plutonium-239 (Pu). The Pu contamination resulted from experiments in which assemblies of Pu were chemically exploded to test for safety from nuclear fission in the event a nuclear weapon is involved in an accident (Gilbert et al. 1988).

Suppose that the mean concentration of Pu in surface soil must be estimated, to prepare for making remedial action decisions in these areas. To do this, we shall use the fact that the radionuclide Americium-241 (Am) is also present in surface soil in these areas. Furthermore, Am has a positive correlation with Pu. That is, if a soil sample has a high concentration of Am, it is also likely to have a high concentration of Pu. Finally, it is known that Am concentrations in surface soil can be approximated at low cost using the FIDLER (Field Instrument for the Detection of Low Energy Radiation), a hand-held gamma detector that gives count-per-minute (cpm) readings of Am.

Suppose that an evaluation of sampling design options has led to a decision to use RSS at one of these NTS testing areas. Suppose further that the required precision of the mean Pu concentration in soil in that area can be achieved if m = 3 soil samples are collected and a measurement of Pu is made on each sample using the "accurate" measurement method. (In actual fact, soil Pu measurements from weapons testing areas on the NTS can vary substantially, so more than 3 samples would usually be required to estimate the mean. We use 3 here for ease of illustration.) Finally, suppose that the ranking of sampling units (field locations within the testing area) will be done using properly calibrated FIDLER detectors that give cpm readings of Am.

For this example, the first three steps of the RSS ranking and soil-sample selection process are:

1. Select a set of m = 3 field locations (sampling units) from the study site using simple random sampling. Call this set S_1. At each of the 3 locations in set S_1, take a FIDLER reading of Am. Rank the 3 locations. That is, assign the rank 1 to the location with the smallest FIDLER reading, assign the rank 2 to the location with the next largest FIDLER reading, and assign the rank 3 to the location with the largest FIDLER reading. Collect a soil sample only at the location that had the ***smallest*** FIDLER reading, that is, at the location given the rank 1. Do ***not*** collect samples at the locations with ranks 2 and 3.

2. Select a second set of m = 3 field locations using simple random sampling. Call this set S_2. Rank the 3 locations in S_2 using FIDLER readings. Collect a soil sample only at the location in set S_2 that has rank 2. Do ***not*** collect soil samples at the locations in S_2 with ranks 1 and 3.

3. Select a third (and final) set of 3 field locations using simple random sampling. Call this set S_3. Rank the 3 locations in S_3 using FIDLER readings. Collect a soil sample only at the location in the set S_3 that has rank 3. Do ***not*** collect soil samples at locations in S_3 with ranks 1 and 2.

At this point, the required m = 3 soil samples have been collected and the ranking and sample selection process ends. The m soil samples are measured for Pu using the "accurate" method as specified in the DQOs. We denote the three Pu measurements by x_1, x_2, and x_3.

The true mean concentration, μ, of Pu for the study area from which the three sets of samples were selected at random is estimated by computing the arithmetic mean ($\bar{x}$) of the 3 Pu measurements:

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i \qquad (1)$$

$$= \frac{1}{3} \sum_{i=1}^{3} x_i$$

The estimated variance of $\bar{x}$, $[s^2(\bar{x})]$, and standard error (SE) are computed using the usual formulas:

$$s^2(\bar{x}) = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{m(m-1)} \qquad (2)$$

$$= \frac{\sum_{i=1}^{3} (x_i - \bar{x})^2}{3(2)}$$

$$SE = [s^2(\bar{x})]^{1/2} \qquad (3)$$

The ranking and soil selection process can be generalized to handle any number of m samples. This is done by generalizing Steps 1 through 3 to Steps 1 through m. At the ith step, the process is

- Select a set, S_i, of m field locations using simple random sampling. Take a FIDLER reading at each of the m locations in set S_i. Rank the m locations in order of increasing FIDLER readings. Collect a soil sample at the location in set S_i that has rank I. Do ***not*** collect a soil sample at any of the other m-1 locations in S_i.

Note that if m samples are needed to meet DQO requirements, RSS requires that m^2 locations be ranked in the process of selecting m locations for sample collection. In the example above, $m^2 = 3^2 = 9$ locations must be ranked in the process of selecting m = 3 soil samples. Similarly, if m = 20 soil samples are needed to meet DQO requirements, then $m^2 = 20^2 = 400$ locations must be ranked in the process of selecting m = 20 soil samples. Clearly, the number of units that must be ranked increases rapidly as m increases. Hence, it is very important to use a low-cost ranking method unless m is very small.

3.0 Numerical Example

For the case of m = 3, suppose that the FIDLER readings and their ranks obtained by applying Steps 1 through 3 are as shown in the table below.

Set S_1		Set S_2		Set S_3	
FIDLER	Rank	FIDLER	Rank	FIDLER	Rank
20,000	1	20,000	1	40,000	1
45,000	2	***90,000***	2	50,000	2
120,000	3	230,000	3	***100,000***	3

Therefore, following the steps in Section 2, in set S_1 a soil sample is collected at the location with FIDLER reading 20,000 cpm, in set S_2 at the location with FIDLER reading 90,000 cpm, and in set S_3 at the location with FIDLER reading 100,000 cpm. Suppose the resulting three Pu measurements for these samples are 12.4, 29.5, and 39.6 nCi/g, respectively. Then, using Equations (1), (2), and (3), we obtain

$$\bar{x} = 27.2 \text{ nCi/g}$$

$$SE = 7.9 \text{ nCi/g}$$

That is, the estimate of the true mean Pu concentration for the site is 27.2 nCi/g and the standard deviation of that estimated mean (i.e., the standard error) is 7.9 nCi/g.

4.0 Replicating the RSS Process

If desired, the basic RSS design as discussed in Section 2 can be replicated r times. When replication is used, RSS involves ranking $n = m^2r$ sampling units and collecting n = mr samples, where $r \geq 1$ and $m \geq 2$. Note that when r > 1, $\bar{x}$, $s^2(\bar{x})$, and SE may be computed using the same formulas [Equations (1), (2), and (3)] as when r = 1, *except* that m is replaced by n = mr wherever m appears in the formulas.

Suppose that applying the DQO process resulted in a decision that 12 samples were needed. Using RSS, 12 samples could be selected in one of two ways. One is to first select 12 sets of 12 sampling units using simple random sampling for each set. Then one sample from each set is selected using the ranking and selection process illustrated in Section 2 for the NTS example. For this RSS design, $(12)^2 = 144$ locations would be ranked to select the 12 locations to be sampled. An alternative RSS design would be to first select 3 sets of 3 sampling units using RSS (as illustrated in Steps 1, 2 and 3 in Section 2), and then to replicate that entire process 4 times. For that RSS design, a total of $3^2 \times 4 = 36$ locations would be ranked, and a soil sample collected from 3 x 4 = 12 of those locations.

The choice between these two options would be based on cost and the method expected to yield the most information (smaller SE). These evaluations would be conducted as part of the DQO process. Some guidance is given in Patil et al. (1994b, p. 177) based on the work of Takahasi and Wakimoto (1968). We find that, ignoring cost considerations, the design m = 12 and r = 1 would result in a smaller SE than if m = 3 and r = 4 were used. However, more ranking is required for the m = 12 design, which could increase cost. Also, for some ranking procedures, ranking may become more difficult as m increases, in which case there is motivation to keep m small. When ranking is done by a concomitant variable, such as FIDLER readings (Section 3), Patil et al. (1994b) indicate on the basis of work by Stokes (1977) that m should be chosen as large as is practical.

5.0 Discussion

Interpretation and advice concerning the use of RSS is provided in this section, where r = 1 without loss of generality.

5.1 Unbiased Estimator of the True Mean

Takahasi and Wakimoto (1968) proved that the arithmetic mean ($\bar{x}$) of RSS sample data [Equation (1)] is an unbiased estimator of the true mean, μ, of the target population. That is, by definition, if the steps in the RSS process, including replications, could be repeated K times (where K is very large), the mean of the K arithmetic means ($\bar{x}$) would equal the true population mean. This desirable property is not unique to RSS. It also applies to the standard computing formulas for estimating the true mean when data are collected using other designs, including simple random sampling.

5.2 Estimating the True Variance of $\bar{x}$

If data are obtained using RSS, Equation (2) is an unbiased estimator of the true variance of $\bar{x}$ only if the number of samples collected for accurate measurements is reasonably large. In particular, if m is small, $s^2(\bar{x})$ computed using Equation (2) will tend on the average to be too large (Stokes 1980). This bias is likely to be unimportant if, say, $m \geq 20$, unless the population is highly skewed. However, this issue needs further study.

5.3 Relative Precision of RSS Compared to Simple Random Sampling

RSS yields more precise statistical estimates than does simple random sampling. The estimated mean ($\bar{x}$) based on m measurements obtained using RSS will be more precise than the estimated mean based on m measurements obtained using simple random sampling. In particular, Takahasi and Wakimoto (1968) determined that when the ranking is perfect, the relative precision (RP) of RSS compared to simple random sampling is between 1 and (m+1)/2; that is,

$$1 \leq RP \leq \frac{m+1}{2} \qquad (4)$$

where

$$RP = \frac{\text{Variance of } \bar{x} \text{ When Data Are Obtained Using Simple Random Sampling}}{\text{Variance of } \bar{x} \text{ When Data Are Obtained Using RSS}}$$

The upper limit [RP = (m+1)/2] is attained only when the underlying distribution of the data is rectangular (i.e., when all values between a lower and upper bound are equally likely to be measured). Hence, for that distribution, almost (m+1)/2 times as many measurements on units selected using simple random sampling are required to equal the precision of the RSS estimator based on m samples (Patil et al. 1994a). For any other specific underlying distribution, the RP will lie between the limits given above in Equation (4). The lower limit (RP = 1) occurs when ranking is of no value, that is, when ranking is no better than assigning ranks at random to the initial set of randomly selected units. This latter result indicates that we never lose any precision by ranking the initial set of randomly selected units, no matter how poor the ranking may be. Of course, the increase in RP is purchased at the cost of ranking the units. Such costs must be considered when deciding whether to use RSS.

Table 1 in Patil et al. (1994b) gives the RP of RSS relative to simple random sampling for various population distributions (including normal, rectangular, exponential, gamma, Weibull, and triangular). The Patil et al. table indicates that RP = 1.914 if the population has a normal distribution with mean 0 and standard deviation 1, and if RSS with m = 3 measurements will be used to compute $\overline{x}$ (as was the case for the NTS example we presented in Section 2). That is, the true variance of $\overline{x}$ when 3 measurements obtained using simple random sampling will be used to compute $\overline{x}$ is almost twice as large as the true variance of $\overline{x}$ when the 3 measurements are obtained using the RSS procedure. If m = 5 measurements will be used, then RP increases to 2.77. Similar results are obtained for all distributions in Table 1 of Patil et al. (1994b). Hence, it appears that in practice it would be wise to use RSS instead of simple random sampling unless the cost of the ranking process used in RSS is prohibitively high.

Patil et al. (1994b) also determined the RP when the population distribution is lognormal. Their Table 2 indicates that the RP decreases as the standard deviation of the logarithms increases, that is, as the skewness or coefficient of variation of the lognormal distribution increases. For example, when m = 3 or 10 and the data have a lognormal distribution with standard deviation, σ (in logarithmic units), they obtained

	σ	0.1	0.5	1.0
m = 3	RP	1.90	1.70	1.34
m = 10	RP	4.73	3.58	2.06

When ranking is imperfect, it is still true that RP ≥ 1. However, as the errors in ranking become large, the RP will move closer to 1. When the ranking variable is a measured quantity (e.g., the FIDLER cpm readings in our NTS example), the RP will depend on how well the ranking variable and the variable of interest (Pu concentrations in the example) are correlated (David and Levine 1972; Stokes 1977; Patil et al. 1994a, 1994b).

5.4 Statistical Methods Available for Use on Data Obtained Using RSS

The objective of collecting environmental data often extends beyond simply estimating the true mean. Other objectives include computing 95% confidence limits for the true mean, estimating trends over time or space, or testing whether a risk-based or background-based cleanup standard has been attained by remedial actions. Many of these alternative applications have not yet been fully developed for RSS. However, methods have recently been developed to 1) test for a difference in the median concentration of two populations using a modified Wilcoxon Rank Sum test (Bohn and Wolfe 1992, 1994), and to 2) estimate the cumulative distribution function (cdf) and a confidence band for the cdf (Stokes and Sager 1988). Bohn and Wolfe (1992) show that RSS can, in some situations, lead to a more powerful Wilcoxon Rank Sum test. Stokes and Sager (1988) show that RSS is more efficient than simple random sampling for estimating the cdf.

Patil et al. (1994b) review the theory, methods, and applications of RSS. They also demonstrate the use of RSS for improving the formation of composite samples. To our knowledge, their article gives the most complete review of RSS currently available.

Patil et al. (1992b) discuss applications of RSS to ecological data analysis. They also examine the relative precision of RSS to simple random sampling when the data follow the lognormal, Poisson, logarithmic, or binomial distribution. Graphs of the relative precision illustrate the superior performance of RSS compared to simple random sampling when data have these distributions.

Patil et al. (1992c) examine the performance of RSS compared to simple random sampling, stratified random sampling, and systematic sampling when the population has a trend. They found that RSS provides a more precise estimator of the true mean than simple random sampling when a trend is present. Also, they found that in some cases, the performance of RSS for estimating a mean was better than if stratified random sampling and systematic sampling were used. Patil et al. (1994a, 1994b) illustrate the effectiveness of RSS relative to simple random sampling by using polycholorinated biphenyls (PCB) data from 19 gas pipeline sites in Pennsylvania.

A trio of recent publications shows how to use RSS to estimate parameters of several probability distributions (Lam et al. 1994; Fei et al. 1994; Sinha et al. 1994).

5.5 DQO Environmental Applications of RSS Data

A possible environmental application of RSS is in deciding how to dispose of sealed drums that contain heterogeneous waste of unknown concentrations. *In situ* detector readings might be used to rank randomly selected sets of drums before a subset of those drums is selected for detailed inspection. Another application might be in deciding which samples archived in storage should be measured for a new pollutant of interest. If aliquots from these stored samples had previously been measured for other pollutants or natural constituents, those data might be useful for ranking randomly selected sets of stored samples in the RSS process.

RSS could also be used to resolve *side issues*. For example, in working through the DQO process, many side issues and secondary decisions may be required before the DQO team can begin work on the primary decision. As the team works through the multiple issues at a site, the cost-effective RSS approach can be used to provide estimates of average concentrations for these secondary or supporting issues. These averages can be used to provide support for saying "yes, this may be a problem area and must be considered further." In the Pu example in Section 2, the primary decision may be whether the soil in an area can be stored in some engineered landfill for a fixed time period. The main issue may have to do with a hazardous component of the waste. However, as a side issue, if the average Pu is greater than some level, the engineered landfill is not even an option for consideration.

5.6 Double Sampling

A sampling design called linear regression double sampling [discussed, e.g., in Cochran (1977), Gilbert and Eberhardt (1976), and Gilbert (1987, Ch. 9)] also makes use of auxiliary data such as FIDLER readings to obtain a better estimate of the true mean or to estimate the true mean with reduced cost. The approach is to estimate the linear relationship (using regression analysis) between two types of measurements made on the same units: measurements made using an inexpensive but fallible method and measurements made using an expensive "accurate" method. The linear regression line and the variability in the data about the line are then used to estimate the mean and its standard error. Double sampling works well if the cost of the inexpensive data is sufficiently low compared to the cost of the "accurate" measurement method and if the positive correlation between the two types of measurements is sufficiently large. Patil et al. (1993) studied the precision of RSS relative to double sampling. They concluded that for the population model they considered, RSS performed approximately the same as double sampling unless the positive correlation was larger than 0.85. In that case, double sampling gives a more precise estimate of the true mean.

6.0 Future Activities

This issue of the *DQO Statistics Bulletin* has indicated that RSS offers considerable potential for obtaining better estimates of the true mean with little, if any, increase in costs or, alternatively, to obtain an equally precise estimate of the mean for less cost. However, the potential for using RSS for other purposes should be explored. RSS needs further evaluation to determine where it may be used to best advantage to reduce costs and improve performance of site management and cleanup.

Some suggested topics for further study are

- developing statistical tests that have improved performance when RSS is used
- developing a procedure for computing confidence limits on the true mean using RSS samples
- incorporating composite sampling with RSS to achieve additional savings in sample analysis costs
- developing a procedure for specifying the optimal number of samples for estimating a mean or percentiles of a distribution or for testing hypotheses.

Additional research topics are discussed by Patil et al. (1992a).

7.0 References

Bohn, L.L., and D.A. Wolfe. 1992. "Nonparametric Two-Sample Procedures for Ranked-Set Samples Data." *Journal of the American Statistical Association* 87:552-561.

Bohn, L.L., and D.A. Wolfe. 1994. "The Effect of Imperfect Judgment Ranking on Properties of Procedures Based on the Ranked-Set Samples Analog of the Mann-Whitney-Wilcoxon Statistic." *Journal of the American Statistical Association* 89:168-176.

Cochran, W.G. 1977. *Sampling Techniques*. 3rd ed. Wiley, New York.

David, H.A., and D.N. Levine. 1972. "Ranked Set Sampling in the Presence of Judgment Error." *Biometrics* 28:553-555.

Fei, H., B.K. Sinha, and Z. Wu. 1994. "Estimation of Parameters in Two-Parameter Weibull and Extreme-Value Distributions Using Ranked Set Sample." *Journal of Statistical Research* 28(1&2):149-161.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* Van Nostrand Reinhold, New York.

Gilbert, R.O., J.H. Shinn, E.H. Essington, E.M. Romney, K.S. Moor, and T.P. O'Farrell. 1988. "Radionuclide Transport from Soil to Air, Native Vegetation, Kangaroo Rats and Grazing Cattle on the Nevada Test Site." *Health Physics* 35:869-887.

Gilbert R.O., and L.L. Eberhardt. 1976. "An Evaluation of Double Sampling for Estimating Plutonium Inventory in Surface Soil." In *Radioecology and Energy Sources,* C.E. Cushing, ed., pp. 157-163, Dowden, Hutchison and Ross, Inc., Stroudsburg, Pennsylvania.

Lam, K., B.K. Sinha, and W. Zhong. 1994. "Estimation of Parameters in a Two-Parameter Exponential Distribution Using Ranked Set Sample." *Annals of Institute of Statistical Mathematics* 46(4):723-736.

McIntyre, G.A. 1952. "A Method for Unbiased Selective Sampling Using Ranked Sets." *Australian Journal of Agricultural Research* 3:385-390.

Patil, G.P., S.D. Gore, and A.K. Sinha. 1994a. "Environmental Chemistry, Statistical Modeling, and Observational Economy,. In *Environmental Statistics, Assessment, and Forecasting,* C.R. Cothern and N. P. Ross, eds., pp. 57-97. Lewis Publishers, Boca Raton, Florida.

Patil, G.P., A.K. Sinha, and C. Taillie. 1994b. "Ranked Set Sampling." In *Handbook of Statistics 12*, G.P. Patil and C.R. Rao, eds., pp. 167-200. North-Holland, New York.

Patil, G.P., A.K. Sinha, and C. Taillie. 1992a. *Ranked Set Sampling Annotated Bibliography*. *Technical Report 91-1201*. Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, Pennsylvania.

Patil, G.P., A.K. Sinha, and C. Taillie. 1992b. *Ranked Set Sampling and Ecological Data Analysis,* Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, Pennsylvania.

Patil, G.P., A.K. Sinha, and C. Taillie. 1992c. *Ranked Set Sampling in the Presence of a Trend on a Site*. Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, Pennsylvania.

Sinha, B.K., Z. Wu, and F. Heliang. 1994. "Estimation of a Gamma Mean Based on a Ranked Set Sample." *Pakistan Journal of Statistics* 10(1)A: 235-249.

Stokes, S.L. 1977. "Ranked Set Sampling With Concomitant Variables." *Communications in Statistics, Part A - Theory and Methods* A6(12): 1207-1211.

Stokes, S.L. 1980. "Estimation of Variance Using Judgment Ordered Ranked Set Samples." *Biometrics* 36:35-42.

Stokes, S.L., and T.W. Sager. 1988. "Characterization of a Ranked-Set Sample With Application to Estimating Distribution Functions." *Journal of the American Statistical Association* 83:374-381.

Takahasi, K., and Wakimoto, K. 1968. "On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering." *Annals of Institute of Statistical Mathematics* 20:1-31.

EPA. 1993. *Data Quality Objectives Process for Superfund*. EPA/540/G-93/071, U.S. Environmental Protection Agency, Office of Emergency and Remedial Response, Washington, D.C.

EPA. 1994. *Guidance for the Data Quality Objectives Process*. EPA QA/G-4, U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, D.C.

Copies of this bulletin are currently distributed to

U.S. Department of Energy

Laboratory Management Division EM-263
19901 Germantown Road
Germantown, MD 20874-1290
D. Bottrell

Office of Environmental Technology
and Waste Management
19901 Germantown Road
Germantown, MD 20874-1290
M. Carter

Environmental Safety and Health EH-232
1000 Independence Avenue SW
Washington, DC 20585
H.R. Peterson, Jr.
A. Wallo, III
R. Dailey

Richland Operations Office
Richland, WA 99352
R.P. Carter
J.M. Clark
J.W. Day
P.K. Clark
J.B. Hall

Nevada Operations Office
P.O. Box 98518
Las Vegas, NV 89193-8518
D. Duncan
D. Hippensteal
S.A. Mellington

Rocky Flats Office
P.O. Box 928
Golden, CO 80402-0928
B.K. Thatcher

U.S. Environmental Protection Agency

401 M Street SW
Washington, DC 20460
J.R. Brown
D. Geuder
W.B. Johnson
B.D. Nussbaum
N.P. Ross
J. Warren

EPA Region IV
960 College Station Road
Athens, GA 30613
W. Cosgrove

P.O. Box 93478
Las Vegas, NV 89193
E.J. Englund
G.T. Flatman

AREAL (MD-75)
Research Triangle Park, NC 27711
L.H. Cox

Region 10
Richland, WA 99352
P.S. Innis

U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Washington, DC 20555
A.M. Huffert

Reynolds Electrical & Engineering Co., Inc.
P.O. Box 98521, M/S 417
Las Vegas, NV 89193-8521
R.R. Kinnison

Sandia National Laboratories
Albuquerque, NM 87185-5800
R.G. Knowlton

Boeing Computer Services Richland
P.O. Box 300, B4-63
Richland, WA 99352
T. LeGore

Desert Research Institute
P.O. Box 19040
Las Vegas, NV 89132-0040
F.L. Miller

Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6367
M.D. Morris

Neptune and Company
P.O. Box 10
Los Alamos, NM 87544
D. Neptune

Penn State University
Department of Statistics
421 Classroom Building
University Park, PA 16802
G.P. Patil

MACTEC

107 Edinboro Lane, Suite C-30
Oak Ridge, TN 37830-7698
K.S. Redus

2309 Renard Place SE, Suite 103
Albuquerque, NM 87106
S. Blacker
K. Hull

639 Cullum
Richland, WA 99352
J.A. Sheriff

Westinghouse Savannah River Company
P.O. Box 616
Aiken, SC 29802
J.D. Heffner
R.C. Tuckfield

Westinghouse Hanford Company
Richland, WA 99352
L. Jensen
D.E. Clark

EG&G Idaho Inc.
P.O. Box 1625
Idaho Falls, ID 83415
C.L. Atwood
W.G. Blackwood

EG&G Rocky Flats Inc.
Statistical Applications
P.O. Box 464
Golden, CO 80402-0464
D.R. Weier

Los Alamos National Laboratory
Los Alamos, NM 87545
R.J. Beckman
K.S. Campbell
E.J. Kelley

Washington State Department of Ecology
P.O. Box 47600
Olympia, WA 98504-7600
N. Blakley
C. Cline
A. Stone

Lawrence Livermore National Laboratory
P.O. Box 808
Livermore, CA 94551-9900
L.R. Anspaugh
S. Blacker

University of Washington
Center for Quantitative Science HR20
3737 15th Avenue, NE
Seattle, WA 98195
L.L. Conquest

Environmetrics & Statistics Ltd.
3000 Highview Drive, #2323
Henderson, NV 89014
C.B. Davis

Martin Marietta Energy Systems, Inc.
P.O. Box 2009
Oak Ridge, TN 37831-6367
D.J. Downing

National Water Research Institute
Box 5050
Burlington, Ontario L7R 4A6 CANADA
A.H. El-Shaarawi

Environmental Measurements Laboratory
376 Hudson Street
New York, NY 10014-3621
C.V. Gogolak

Electric Power Research Institute
P.O. Box 10412
Palo Alto, CA 94304
R. Goldstein

SENES Oak Ridge Inc.
677 Emory Valley Road, Suite C
Oak Ridge, TN 37830
F.O. Hoffman

The University of Maryland
Baltimore County Campus
Department of Mathematics and Statistics
Baltimore, MD 21228-5398
B.K. Sinha

PPC
1776 Fowler, Suite 15
Richland, WA 99352
J.D. Smyth

Splitstone & Associates
4556 William Penn Highway, Number 110
Murrysville, PA 15668
D.E. Splitstone

Bechtel Hanford Inc.
450 Hills Street
Richland, WA 99352
S.C. Tindall

Pacific Northwest Laboratory
P.O. Box 999
Richland, WA 99352
D.J. Bates
G. Chen
P.G. Doctor
F. GAO
R.O. Gilbert
N.L. Hassig
A.M. Liebetrau
R.C. Locke
J.A. Mahaffey
P.D. Meyer
B.L. Mohler
C.J. Murray
N.K. Neerchal
R.F. O'Brien
T.L. Page
B.A. Pulsipher
R.D. Quinn
R.E. Rossi
S.C. Slate
W.G. Stillwell

Credits for this issue

The help of Nancy Hassig, Derrick (Rick) Bates, Brent Pulsipher, Bob O'Brien, Nagaraj Neerchal, Tim LeGore, and Dave Bottrell for reviewing drafts of this bulletin and for helpful discussions of possible applications of RSS is gratefully acknowledged. Their suggestions and insight resulted in substantial improvements to the text. Special thanks are also due to Andrea Currie for her help in developing the design and format of the bulletin and for editorial and production assistance.