Environmental and Ecological Statistics **8,** 151–162, 2001

Identifying the largest individual sample value from a two-way composite sampling design

S.D. GORE, G.P. PATIL* and C. TAILLIE

Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA 16802

Received August 1997; Revised November 2000

Composite sampling techniques for identifying the largest individual sample value seem to be cost effective when the composite samples are internally homogeneous. However, since it is not always possible to form homogeneous composite samples, these methods can lead to higher costs than expected. In this paper we propose a two-way composite sampling design as a way to improve on the cost effectiveness of the methods available to identify the largest individual sample value.

Keywords: classification, composite sampling, composite sampling design, estimation, extreme values, locally sequential sweep-out method, globally sequential sweep-out method, sweep-out method

1. Introduction

Composite sampling techniques were initially developed for classifying sampling units without requiring separate measurements on all individual sampling units (Dorfman, 1943). Later developments extended the application of composite sampling to estimation of the population mean (see, for instance, Duncan, 1962; Brown and Fisher, 1972; Rohde, 1976; Elder, Thompson, and Myers, 1980), the population variance and its components (Brumelle, Nemetz, and Casey, 1984), testing of hypotheses (Ryti, Neptune, and Groskinsky, 1992) , sample size determination (see Rohlf *et al*., 1991), and so on. In spite of this development, a major limitation of composite sampling has been the loss of information on individual sample values. This limitation becomes more apparent when the objectives of a study include identification of ''hotspots'', i.e., extremely large sample values. The problem of recovering the largest individual sample value using the composite sample data has been addressed more recently. For instance, Casey, Nemetz, and Uyeno

Prepared with partial support from the Statistical Analysis and Computing Branch, Environmental Statistics and Information Division, Office of Policy, Planning, and Evaluation, United States Environmental Protection Agency, Washington, DC under a Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

*Author for correspondence

1352-8505 © 2001 Kluwer Academic Publishers

(1985) give a method to predict the largest individual sample value by making measurements on individual samples in the composite sample having the largest measured value. Gore and Patil (1994) give a ''sweep-out'' method to identify the largest individual sample value, where the number of measurements on individual samples is not fixed, but the largest individual sample value is identified with certainty. Gore, Patil, and Taillie (1996) develop certain modifications of the sweep-out method with a view to reduce the number of measurements needed on individual samples. Aragon, Gore, and Patil (1994) introduce and illustrate a concept of a two-way compositing design. Patil and Taillie (2001) investigate statistical aspects of the two-way compositing design. They also provide a geostatistical justification to the intuitive logic for an efficient search of the largest individual sample value in a two-way compositing design. In this paper, we investigate a two-way composite sampling design for identification of the largest individual sample value, using a suitable two-way version of the sweep-out method.

Recall the sweep-out method of Gore and Patil (1994). This procedure assumes that composite sample measurements are made without error. The sweep-out method also assumes that every individual sampling unit is available for measurement at any stage of the procedure. With these assumptions and with the composite sample values that are available, the sweep-out method proceeds as follows. The composite samples are arranged in the descending order of their values, so that the leading composite sample in this arrangement becomes the candidate composite sample having the largest measured value. This composite sample is therefore labelled for retesting, i.e., for making measurements on its constituent individual sampling units. Measurements are made on the individual samples in the composite selected for retesting and the largest of these individual sample values is defined as the composite sample maximum. When this value is obtained for the first time, it is also defined as the record value. At subsequent occasions, the larger of the record value and the composite sample maximum is defined as the updated record value. After determining the record value, all the unretested composite samples that have totals exceeding the record value are arranged in the descending order of their values. The leading composite sample in this arrangement is then selected for retesting. The procedure continues as long as the list of composite samples having totals exceeding the record value is not empty. At the end of the procedure, the record value is declared as the largest individual sample value.

The locally sequential sweep-out method of Gore, Patil, and Taillie (1996) improves on the above sweep-out method by eliminating unnecessary measurements on individual samples in those composite samples that are selected for retesting. The globally sequential sweep-out method of Gore, Patil, and Taillie (1996) attempts to reduce the number of measurements on individual samples by selecting an individual sample for measurement from among all individual samples that have not been subjected to measurement already. In the sequential sweep-out methods described above, an individual sample is selected randomly from a composite sample that is selected for retesting. In absence of any information regarding the magnitude of the individual sample values, there is no alternative way of selecting individual samples. A two-way compositing design is expected to provide some information on individual sample values in a composite sample that is selected for retesting. In this paper, we introduce a two-way version of the sweep-out method to identify the largest individual sample value with a two-way composite sampling design.

We use notation consistent with that of Gore and Patil (1994) and Patil and Taillie

(2001). In Section 2, we describe the proposed method and give an algorithm for its implementation. As analytical results do not seem to be feasible, simulations are used to assess the proposed method. Section 3 provides an application of the algorithm to a 4×4 compositing design. Then in the final section, we discuss a real example based on the cadmium concentrations data from surface soil samples at a hazardous waste site and demonstrate the relative cost advantages of the two-way compositing.

2. A two-way compositing design

A two-way composite sampling design begins by forming a rectangular or square array of sampling units. Corresponding to every row, a row-composite is formed from the individual samples in that row. Individual samples in the columns are likewise used to form column-composites. This two-way compositing design thus assumes that every individual sample contributes to a row-composite and to a column-composite. Measurements are then made on all the row-composites as well as on all the column-composites.

Suppose there are r rows and c columns in the array of the individual samples. For $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$, let X_{ij} denote the value of the individual sample that contributes to the i-th row-composite and the j-th column-composite, $Y_{i\cdot}$ denote the value of the i-th row-composite, and $Y_{\cdot j}$ denote the value of the j-th column-composite. Now arrange the individual samples in the array in such a way that the row-composites as well the column-composites are in descending orders of their respective values. Let $Y_{(i)\cdot}$ denote the value for the i-th ordered row-composite, $i = 1, 2, \ldots, r$, and $Y_{\cdot(j)}$ denote the value for the j-th ordered column-composite, $j = 1, 2, \ldots, c$. That is,

$$Y_{(1)\cdot} \geq Y_{(2)\cdot} \geq \cdots \geq Y_{(r)\cdot},$$

and

$$Y_{\cdot(1)} \geq Y_{\cdot(2)} \geq \cdots \geq Y_{\cdot(c)}.$$

For $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$, let $X_{[i][j]}$ denote the individual sample value for the sample that belongs to the i-th ordered row-composite and the j-th ordered column-composite. Similarly, for $i = 1, 2, \ldots, r$, let $W_{[i]\cdot}$ denote the composite sample total for the i-th ordered composite, and for $j = 1, 2, \ldots, c$, let $W_{\cdot[j]}$ denote the composite sample total for the j-th column-composite. The following table explains the arrangement of individual samples so that the row-composite values as well as the column-composite values are in the descending order.

	Column					
Row	1	2	$\cdots$	c	*Value*	*Total*
1	$X_{[1][1]}$	$X_{[1][2]}$	$\cdots$	$X_{[1][c]}$	$Y_{(1)\cdot}$	$W_{[1]\cdot}$
2	$X_{[2][1]}$	$X_{[2][2]}$	$\cdots$	$X_{[2][c]}$	$Y_{(2)\cdot}$	$W_{[2]\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	$X_{[r][1]}$	$X_{[r][2]}$	$\cdots$	$X_{[r][c]}$	$Y_{(r)\cdot}$	$W_{[r]\cdot}$
Value	$Y_{\cdot(1)}$	$Y_{\cdot(2)}$	$\cdots$	$Y_{\cdot(c)}$		
Total	$W_{\cdot[1]}$	$W_{\cdot[2]}$	$\cdots$	$W_{\cdot[c]}$		

With the composite sample values, and hence also the composite sample totals available for the r row-composites and the c column-composites, we suggest the following two-way sweep-out method.

Begin by making a measurement to obtain the value $X_{[1][1]}$ and denote it by R to indicate that it is the record value among the individual sample measurements. Subtract the value of $X_{[1][1]}$ from $W_{[1]\cdot}$ and $W_{\cdot[1]}$. Update the composite sample sizes and composite sample values in the first row and the first column. Replace $X_{[1][1]}$ with a 0 in the table, so that the cell representing this individual sample will not be selected again for making measurement. Now, rearrange the individual samples so that the updated composite sample values are again in the descending order along rows as well as along columns. Define i as the smallest row index corresponding to a row-composite having the composite sample total exceeding the value of R. Then define j as the smallest column index corresponding to a column-composite having the composite sample total exceeding the value of R and having the individual sample representing the cell (i,j) still unmeasured. Make a measurement on the individual sample representing the cell (i,j) and denote it by $X_{[i][j]}$. Replace R with the larger of R and $X_{[i][j]}$, subtract $X_{[i][j]}$ from $W_{[i]\cdot}$ and $W_{\cdot[j]}$, decrement $k_{[i]\cdot}$ and $k_{\cdot[j]}$ by 1 each, compute the updated composite sample values for the row-composite i and the column-composite j, and replace $X_{[i][j]}$ by a 0 in the cell (i,j). At this stage, rearrange the individual samples so that the updated composite sample values are again in the descending order along rows as well as along columns. Repeat the above procedure as long as there is at least one cell that has not been subjected to measurement, and belongs to a row-composite and to a column-composite that have composite sample totals exceeding the value of R. When the procedure terminates because there is no cell satisfying this condition, the value of R is declared as the largest individual sample value.

We now give a formal algorithm for the two-way sweep-out method described above.

1. Define $R = 0$ and $K = 0$.
2. Arrange the $r \times c$ individual samples in such a way that the row-composites and the column-composites have their values in the descending order.
3. Define
$$\mathcal{U}_R = \{h : 1 \leq h \leq r, W_{[h]\cdot} > R\}$$
$$\mathcal{U}_C = \{h : 1 \leq h \leq c, W_{\cdot[h]} > R\}.$$
4. If $\mathcal{U}_R$ or $\mathcal{U}_C$ is empty, then stop the search and declare R as the largest individual sample value.
5. Define $i = \min\{h : h \in \mathcal{U}_R\}$.
6. Define $j = \min\{h : h \in \mathcal{U}_C\}$.
7. If the individual sample representing the cell (i,j) has not been measured so far, then follow Step 10.
8. If $Y_{[i]\cdot} < Y_{\cdot[j]}$, then define i as the next element in $\mathcal{U}_R$, and follow Step 7
9. If $Y_{[i]\cdot} \geq Y_{\cdot[j]}$, then define j as the next element in $\mathcal{U}_C$, and follow Step 7.
10. If $k_{[i]\cdot} > 1$ and $k_{\cdot[j]} > 1$, then follow Step 13.
11. If $k_{[i]\cdot} = 1$, then define $X_{[i][j]} = Y_{[i]\cdot}$ and follow Step 14.
12. If $k_{\cdot[j]} = 1$, then define $X_{[i][j]} = Y_{\cdot[j]}$ and follow Step 14.
13. Make a measurement on the individual sample representing the cell (i,j) to obtain the value of $X_{[i][j]}$. Increment K by 1.

14. Update $R \leftarrow \max\{R, X_{[i][j]}\}$.
15. Update the i-th row-composite: $W_{[i]\cdot} \leftarrow W_{[i]\cdot} - X_{[i][j]}$, $k_{[i]\cdot} \leftarrow k_{[i]\cdot} - 1$, $Y_{[i]\cdot} = W_{[i]\cdot}/k_{[i]\cdot}$. Similarly, update the j-th column-composite: $W_{\cdot[j]} \leftarrow W_{\cdot[j]} - X_{[i][j]}$, $k_{\cdot[j]} \leftarrow k_{\cdot[j]} - 1$, $Y_{\cdot[j]} = W_{\cdot[j]}/k_{\cdot[j]}$. Write 0 in the cell (i, j) and follow Step 2.

3. Illustration of the algorithm

In this section, we illustrate the algorithm of the preceding section with artificial data. We consider a two-way compositing design for 16 individual samples arranged in a square of 4 rows and 4 columns. The individual sample values are as follows.

66	18	140	61
177	45	23	69
82	109	48	40
77	167	36	47

The algorithm given in the preceding section is then applied to the above data. The steps of the algorithm in this application are given below. The leading numeral indicates the step of the algorithm.

1. Define $R = 0$ and $K = 0$.
2. When row-composites and column-composites are formed and the 16 individual samples arranged so that the composite sample values are in the descending order, we have the following table.

					$Y_{[i]\cdot}$	$W_{[i]\cdot}$
	77	167	36	47	82	327
	177	45	23	69	79	314
	66	18	140	61	71	285
	82	109	48	40	70	279
$Y_{\cdot[j]}$	101	85	62	54		
$W_{\cdot[j]}$	402	339	247	217		

3. Define $\mathcal{U}_R = \{1, 2, 3, 4\}$ and $\mathcal{U}_C = \{1, 2, 3, 4\}$.
4. Since $\mathcal{U}_R$ and $\mathcal{U}_C$ are both non-empty, follow Step 5.
5. Set $i = 1$.
6. Set $j = 1$.
7. Since the individual sample representing the cell $(1, 1)$ has not been measured, follow Step 10.
10. Since $k_{[1]\cdot} = 4 > 1$ and $k_{\cdot[1]} = 4 > 1$, follow Step 13.
13. Make a measurement on the individual sample representing the cell $(1, 1)$ to obtain $X_{[1][1]} = 77$.
14. $R = 77$, $K = 1$, and $X_{[1][1]} = 0$.
15. $W_{[1]\cdot} = 250$, $k_{[1]\cdot} = 3$, $Y_{[1]\cdot} = 83$, and $W_{\cdot[1]} = 325$, $k_{\cdot[1]} = 3$, $Y_{\cdot[1]} = 108$.

2. Arranging the individual samples in order that the row-composites and the column-composites have their values in the descending order, we get the following table.

					$Y_{[i]\cdot}$	$W_{[i]\cdot}$
	0	167	36	47	83	250
	177	45	23	69	78	314
	66	18	140	61	71	285
	82	109	48	40	70	279
$Y_{\cdot[j]}$	108	85	62	54		
$W_{\cdot[j]}$	325	339	247	217		

3. $\mathscr{U}_R = \{1, 2, 3, 4\}$ and $\mathscr{U}_C = \{1, 2, 3, 4\}$.
4. Since $\mathscr{U}_R$ and $\mathscr{U}_C$ are both non-empty, follow Step 5.
5. Set $i = 1$.
6. Set $j = 1$.
7. Since the individual sample representing the cell $(1, 1)$ has already been measured, follow Step 8.
8. Since the condition $Y_{[i]\cdot} \geq Y_{\cdot[j]}$ is not satisfied, follow Step 9.
9. Since $Y_{[i]\cdot} < Y_{\cdot[j]}$, set $i = 2$ and follow Step 7.
7. Since the individual sample representing the cell $(2, 1)$ has not been measured, follow Step 10.
10. Since $k_{[2]\cdot} = 4 > 1$ and $k_{\cdot[1]} = 3 > 1$, follow Step 13.
13. Make a measurement on the individual sample representing the cell $(2, 1)$ to obtain $X_{[2][1]} = 177$.
14. $R = 177$, $K = 2$, and $X_{[2][1]} = 0$.
2. Arranging the individual samples in order that the row-composites and the column-composites have their values in the descending order, we get the following table.

					$Y_{[i]\cdot}$	$W_{[i]\cdot}$
	167	0	36	47	83	250
	18	66	140	61	71	285
	109	82	48	40	70	279
	45	0	23	69	46	137
$Y_{\cdot[j]}$	85	74	62	54		
$W_{\cdot[j]}$	339	148	247	217		

3. $\mathscr{U}_R = \{1, 2, 3\}$ and $\mathscr{U}_C = \{1, 3, 4\}$.
4. Since $\mathscr{U}_R$ and $\mathscr{U}_C$ are both non-empty, follow Step 5.
5. Set $i = 1$.
6. Set $j = 1$.
7. Since the individual sample representing the cell $(1, 1)$ has not been measured, follow Step 10.
10. Since $k_{[1]\cdot} = 3 > 1$ and $k_{\cdot[1]} = 4 > 1$, follow Step 13.
13. Make a measurement on the individual sample representing the cell $(2, 1)$ to obtain $X_{[2][1]} = 167$.
14. $R = 177$, $K = 3$, and $X_{[1][1]} = 0$.

2. Arranging the individual samples in order that the row-composites and the column-composites have their values in the descending order, we get the following table.

					$Y_{[i]\cdot}$	$W_{[i]\cdot}$
	66	140	18	61	71	285
	82	48	109	40	70	279
	0	23	45	69	46	137
	0	36	0	47	42	83
$Y_{\cdot[j]}$	74	62	57	54		
$W_{\cdot[j]}$	148	247	172	217		

3. $\mathscr{U}_R = \{1, 2\}$ and $\mathscr{U}_C = \{2, 4\}$.
4. Since $\mathscr{U}_R$ and $\mathscr{U}_C$ are both non-empty, follow Step 5.
5. Set $i = 1$.
6. Set $j = 2$.
7. Since the individual sample representing the cell $(1, 2)$ has not been measured, follow Step 10.
10. Since $k_{[1]\cdot} = 4 > 1$ and $k_{\cdot[1]} = 4 > 1$, follow Step 13.
13. Make a measurement on the individual sample representing the cell $(1, 2)$ to obtain $X_{[1][2]} = 140$.
14. $R = 177$, $K = 4$, and $X_{[1][2]} = 0$.
2. Arranging the individual samples in order that the row-composites and the column-composites have their values in the descending order, we get the following table.

					$Y_{[i]\cdot}$	$W_{[i]\cdot}$
	82	109	40	48	70	279
	66	18	61	0	48	145
	0	45	69	23	46	137
	0	0	47	36	42	83
$Y_{\cdot[j]}$	74	57	54	36		
$W_{\cdot[j]}$	148	172	217	107		

3. $\mathscr{U}_R = \{1\}$ and $\mathscr{U}_C = \{3\}$.
4. Since $\mathscr{U}_R$ and $\mathscr{U}_C$ are both non-empty, follow Step 5.
5. Set $i = 1$.
6. Set $j = 3$.
7. Since the individual sample representing the cell $(1, 3)$ has not been measured, follow Step 10.
10. Since $k_{[1]\cdot} = 4 > 1$ and $k_{\cdot[3]} = 4 > 1$, follow Step 13.
13. Make a measurement on the individual sample representing the cell $(1, 3)$ to obtain $X_{[1][3]} = 40$.
14. $R = 177$, $K = 5$, and $X_{[1][3]} = 0$.
2. Arranging the individual samples in order that the row-composites and the column-composites have their values in the descending order, we get the following table.

					$Y_{[i]\cdot}$	$W_{[i]\cdot}$
	82	0	109	48	80	239
	66	61	18	0	48	145
	0	69	45	23	46	137
	0	47	0	36	42	83
$Y_{\cdot[j]}$	74	59	57	36		
$W_{\cdot[j]}$	148	177	172	107		

3. $\mathscr{U}_R = \{1\}$ and $\mathscr{U}_C = \phi$.
4. Since $\mathscr{U}_C$ is empty, declare $R = 177$ as the largest individual sample value and stop the search.

Note that the algorithm requires a total of $K = 6$ individual sample measurements to identify the largest individual sample value. Of these six individual sample measurements, two are required to encounter the largest individual sample value 177 and four more measurements are required to confirm that this indeed is the largest individual sample value.

4. Illustrative application

In this section, we illustrate the two-way sweep-out method with an example based on field data.

The data consist of Cadmium concentrations in surface soil samples at the Palmerton Superfund site. More information on the Palmerton site and the sampling activities at the site are available in Bolgiano, Patil, and Taillie (1990). From the square sampling grid at the site, a square of 24 rows and 24 columns is selected for illustration. Table 1 gives the individual sample values, composite sample sizes, composite sample totals, and composite sample values (averages).

Note that the individual sample values are arranged in such a way that the row-composite values and column-composite values are in the descending order. In this case, the largest individual sample value of 297.5 is encountered on the third selected individual sample, but is confirmed only after a total of 25 individual sample values are identified. These 25 individual sample values are given below in order:

Step	*X*	*Step*	*X*	*Step*	*X*
1	266	10	177	19	46.9
2	103	11	21.55	20	80.7
3	279.5	12	140	21	39.8
4	158	13	140	22	90.5
5	115	14	48.1	23	111
6	179	15	109	24	62.7
7	158.25	16	69.45	25	20
8	76.58	17	18.6		
9	167	18	109		

The relative cost of the two-way sweep-out method as against exhaustive testing is then RC $= \frac{48+25}{309} = \frac{73}{309} = 0.2362$.

Table 1. Two-way composite sampling design for the cadmium data at the Palmerton superfund site. An asterisk denotes a grid point that is not included in the sample.

	Column						
	1	2	3	4	5	6	7
Row							
1	266.0	103.0	279.5	*	*	*	134.0
2	*	*	*	*	*	*	158.0
3	*	115.0	*	*	*	167.0	109.0
4	*	86.7	*	*	*	76.5	81.8
5	21.6	58.7	69.5	179.0	158.25	36.4	48.1
6	*	*	*	*	53.79	*	74.1
7	108.0	*	*	*	*	*	*
8	45.6	*	62.2	*	140.00	46.8	39.8
9	66.7	*	36.5	86.4	*	23.4	*
10	17.0	9.5	10.1	*	65.00	*	62.7
11	26.5	*	76.1	*	24.55	*	20.0
12	*	51.9	*	30.3	18.10	*	15.8
13	*	*	27.3	*	*	10.6	6.9
14	*	*	*	24.9	*	*	6.2
15	*	14.7	*	*	*	*	15.5
16	*	*	*	*	*	*	*
17	*	*	*	7.0	7.16	*	*
18	5.5	*	6.0	*	6.69	14.1	15.9
19	*	*	*	*	*	*	17.5
20	*	*	3.7	*	3.64	9.7	5.5
21	*	*	9.4	*	*	*	*
22	*	*	*	*	*	*	7.7
23	*	*	*	2.8	*	*	7.3
24	*	*	*	*	*	*	2.8
$k_{\cdot j}$	8	7	10	6	9	8	19
$W_{\cdot j}$	556.80	439.5	580.2	330.5	477.2	384.6	828.7
$Y_{\cdot j}$	69.6	62.8	58.0	55.1	53.0	48.1	43.6

	Column						
	8	9	10	11	12	13	14
Row							
1	*	*	*	*	*	*	*
2	*	105.0	*	*	68.6	*	102.3
3	45.1	*	18.2	*	*	37.2	*
4	177.0	*	66.3	*	*	*	*
5	22.5	57.7	140.0	*	18.6	27.2	72.5
6	*	*	*	*	*	*	*
7	*	*	*	*	14.2	*	*
8	68.7	*	60.5	80.7	90.5	72.8	*
9	10.4	*	*	70.5	42.7	17.9	*
10	33.1	18.2	31.6	*	32.8	*	6.7
11	11.4	*	*	24.5	31.3	*	*

Table 1. Continued.

	Column						
	8	9	10	11	12	13	14
Row							
12	*	12.4	13.2	*	9.0	13.0	24.4
13	*	*	9.2	68.7	*	*	11.0
14	*	*	*	*	*	13.6	*
15	17.2	*	*	*	*	*	*
16	5.7	*	*	11.8	14.5	*	*
17	*	29.3	17.0	6.7	*	*	14.7
18	14.1	*	21.0	*	8.1	*	1.3
19	*	*	*	10.2	*	*	*
20	*	10.6	6.7	34.2	3.8	*	4.0
21	*	*	*	6.5	*	*	*
22	*	*	*	*	*	*	*
23	*	*	*	*	*	*	*
24	*	*	*	*	*	*	*
$k_{\cdot j}$	10	6	10	9	11	6	8
$W_{\cdot j}$	405.2	233.2	383.6	313.9	334.1	181.7	236.8
$Y_{\cdot j}$	40.5	38.9	38.4	34.9	30.4	30.3	29.6

	Column						
	15	16	17	18	19	20	21
Row							
1	*	*	*	*	*	*	45.2
2	*	*	*	*	*	*	*
3	*	*	*	60.3	111.0	*	*
4	*	*	*	*	*	*	6.4
5	*	109.0	123.0	38.6	46.9	130.0	49.7
6	*	*	*	*	*	*	*
7	*	*	*	*	*	*	*
8	121.0	*	*	*	42.8	49.7	23.8
9	25.4	16.7	4.0	6.4	8.6	*	*
10	51.3	57.1	*	*	*	*	12.4
11	21.1	30.7	14.7	12.0	45.8	25.8	*
12	*	*	6.9	8.4	21.2	19.0	*
13	*	*	6.5	27.8	16.0	10.5	6.4
14	*	18.8	*	*	5.5	*	*
15	*	*	5.9	*	*	*	*
16	9.9	19.5	*	*	12.3	5.2	*
17	5.0	8.8	*	*	*	5.5	*
18	*	*	*	*	*	*	*
19	5.7	8.2	*	*	6.4	3.7	*
20	5.0	*	*	*	*	3.7	*
21	11.8	9.7	*	*	5.4	5.7	*
22	*	*	*	*	2.7	*	*
23	*	3.2	*	*	7.0	5.1	*

Table 1. Continued.

	Column						
	15	16	17	18	19	20	21
Row							
24	*	*	*	*	*	*	*
$k_{\cdot j}$	9	10	6	6	13	11	6
$W_{\cdot j}$	256.30	281.8	161.0	153.6	331.6	263.8	143.9
$Y_{\cdot j}$	28.5	28.2	26.8	25.6	25.5	24.0	24.0

	Column					
	22	23	24	$k_{\cdot j}$	$W_{\cdot j}$	$Y_{\cdot j}$
Row						
1	*	35.9	*	6	863.6	143.9
2	*	*	*	4	433.9	108.5
3	47.400	16.8	*	10	727.0	72.7
4	*	8.8	*	7	503.6	71.9
5	45.200	32.3	*	21	1484.7	70.7
6	*	*	*	2	127.9	63.9
7	*	*	*	2	122.2	61.1
8	12.000	5.8	*	16	962.7	60.2
9	*	7.4	3.3	15	426.4	28.4
10	12.800	7.4	6.9	16	434.5	27.2
11	13.300	*	2.5	15	380.2	25.3
12	11.600	*	17.8	15	273.2	18.2
13	13.800	*	*	12	214.5	17.9
14	4.200	*	26.2	7	99.4	14.2
15	*	*	*	4	53.3	13.3
16	29.500	*	4.5	9	113.0	12.6
17	*	*	*	9	101.1	11.2
18	*	*	*	9	92.7	10.3
19	*	*	*	6	51.7	8.6
20	*	*	*	11	90.4	8.2
21	3.630	*	6.2	8	58.4	7.3
22	*	*	*	2	10.5	5.2
23	*	*	*	5	25.6	5.1
24	*	*	*	1	2.8	2.8
$k_{\cdot j}$	10	7	7			
$W_{\cdot j}$	193.4	114.3	67.5			
$Y_{\cdot j}$	19.3	16.3	9.64			

References

Aragon, A.D., Gore, S.D., and Patil, G.P. (1994) Environmental sampling, observational economy, and extreme values. *Parisankhyan Samikkha*, **1**, 71–81.

Bolgiano, N.C., Patil, G.P., and Taillie, C. (1990) Spatial statistics, composite sampling, and related issues in site characteristics with two examples, in *Proceedings of the Workshop on Superfund Hazardous Waste: Statistical Issues in Characterizing a Site, Protocols, Tools, and Research Needs*, H. Lacayo, R.J. Nadeau, G.P. Patil, and L. Zaragoza (eds.), 79–114.

Brown, G.H. and Fisher, N.I. (1972) Subsampling a mixture of sampled material. *Technometrics*, **14**, 663–68.

Brumelle, S., Nemetz, P., and Casey, D. (1984) Estimating means and variances: the comparative efficiency of composite sample grab samples. *Environmental Monitoring and Assessment*, **4**, 81–4.

Casey, D., Nemetz, P.N., and Uyeno, D. (1985) Efficient search procedures for extreme pollutant values. *Environmental Monitoring and Assessment*, **5**, 165–76.

Dorfman, R. (1943) The detection of defective members of large populations. *Ann. Statist.*, **14**, 436–40.

Duncan, A.J. (1962) Bulk sampling: problems and lines of attack. *Technometrics*, **4**, 319–44.

Elder, R.S., Thompson, W.O., and Myers, R.H. (1980) Properties of composite sampling procedures. *Technometrics*, **22**, 179–86.

Gore, S.D. and Patil, G.P. (1994) Identifying extremely large values with composite sample data. (With discussion by John Warren, Henry Kahn and Katherine Campbell, and their rejoinder.) *Environmental and Ecological Statistics*, **1**(3), 227–45.

Gore, S.D., Patil, G.P., and Taillie, C. (1996) Identification of the largest individual sample value using composite sample data and certain modifications of the sweep-out method. *Environmental and Ecological Statistics*, **3**(3), 219–34.

Patil, G.P. and Taillie, C. (2001) Use of best linear unbiased prediction for hot spot identification in two-way compositing. *Environmental and Ecological Statistics* (this issue).

Rohde, C.A. (1976) Composite sampling. *Biometrics*, **32**, 273–82.

Rohlf, F.J., Akcakaya, H.R., and Ferraro, S.P. (1991) *Optimizing composite sampling protocols*, Technical Report prepared for US EPA, Applied Biomathematics, 100 North Country Road, Setauket, NY.

Ryti, R.T., Neptune, D., and Groskinsky, B. (1992) Superfund soil cleanup. *Environmental Testing and Analysis*, **1**, 26–31.

Biographical sketches

S.D. Gore is an Associate Professor in the Department of Statistics and the School of Environmental Sciences, University of Poona, Pune, India.

G.P. Patil is Distinguished Professor of Mathematical Statistics and Director of the Center for Statistical Ecology and Environmental Statistics in the Department of Statistics at the Pennsylvania State University.

Dr C. Taillie is Senior Research Associate in the Center for Statistical Ecology and Environmental Statistics in the Department of Statistics at the Pennsylvania State University.