L. Sayn-Wittgenstein*
A. H. Aldred
Canadian Forestry Service
Ottawa, Ont., Canada

Error Analysis by Computer Simulation

Realistic estimates of the separate errors can be entered into the simulation process and an estimate of a survey's accuracy can be obtained.

Introduction

By reference to an example, this report shows the striking advantages of computer simulation in analysing the complex interactions of errors that occur in many operations involving aerial photographs.

Computer simulation is not new in analytic photogrammetry as seen in recent studies (Whiteside and Lipski, 1968: Wolters, 1968). The full potential of simulation does not, however, seem to be exploited; much can be gained by extending its use to the study of the effects of such errors as sampling errors, errors in the identification of objects, blunders in the recording or transfer of data, *random* measurement errors, and systematic errors. However, before discussing simulation, two other approaches to determine the accuracy of a procedure will be mentioned.

Abstract: *Computer simulation is a most effective method with which to study the effect of interacting errors in procedures where errors of widely different characteristics combine to affect the final result.*

Experimentation

This is a well-established and reliable means of testing the accuracy of a method; it could involve the comparison of results with known true values or with reliable values obtained by an established method. Should this course be impractical, an experiment could be designed, based upon repeated measurements, to obtain an impression of consistency which in turn would give some indication of reliability.

One of the common faults of experimentation is that it tends to be expensive and time-consuming. Another limitation is the restriction to currently feasible experimental procedures.

Analytic Methods

This approach starts with a mathematical model of a real procedure and of the errors expected at various stages. Then, either the influence of individual errors is isolated by differentiation, or the rules on standard errors of functions are applied, which are clearly stated by Yates (1960, p. 196).

In general, this is an effective, proven technique for which many examples are available; of direct relevance to the example to be discussed are papers by Schut and van Wijk (1965), Aldred (1967) and Gerrard (1969).

In taking derivatives and summing variances, however, the approach can go only so far, because it is soon forced towards unrealistic assumptions or is simply confined to dealing with only a few sources of error. The reason is that the errors committed in even simple operations tend to have widely different characteristics that lead to complex interactions which may be too difficult for analytic treatment. The complexity of patterns of error propagation will be illustrated in this report.

Computer Simulation

A third approach, simulation, has much in common with experimentation, but it in-

* Forest Management Institute, Department of Fisheries and Forestry. The first-named is Associate Director and the latter a Research Scientist.

TABLE 1. A SIMPLE ERROR SIMULATION PROBLEM

Building No.	*True Height* (B)	*Measurement Error* (e)	*Measured Height* (B^*)
1	36.0	+ 2.0	38.0
2	120.2	−11.6	108.6
3	180.6	+ 0.1	180.7
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
50	87.5	0.0	87.5

volves no actual measurements; the entire measurement process is imitated, usually in a computer. It involves, first the acceptance of a mathematical description (a *model*) of the objects to be measured; the second step is the simulation of the measurement procedures.

The following over-simplified problem illustrates the simulation procedure: if the height of buildings in a city is measured with an instrument that is subject to random, normally distributed errors with a mean of zero ($m=0$) and a standard deviation of five feet ($s=5$ ft.), how far is the estimate of average building height likely to differ from the true average after measurement of 50 buildings?

The first step is to accept a model of the distribution of building heights in the city, representative of the general problem. This might be done by obtaining data from previous measurements, or by writing down values that are believed to be realistic; the use of artificially generated data is common in simulation studies. The result could be the values B shown in column 2 of Table 1. It is important to realize that these values need not be actual values; they are merely realistic values that are assumed to be representative.

The next stage, the simulation of the measurement process, is accomplished by adding positive or negative *measurement errors* to these *true* values B. As it is specified that these errors are normally distributed with $m=0$, $s=5$ ft., a series of pseudo-random normal deviates e would be generated with these characteristics (Col. 3, Table 1) and to them would be added the values B to obtain the *measured* quantities:

$$B + e = B^*.$$

It is now a simple matter to calculate the average of 50 values of B and B^* and to compare the results.

Of course, a different array of errors will be produced in each simulated set of measurements, just as in actual experiments. The simulation process may therefore have to be repeated until a reliable answer is reached, just as an experiment may have to be repeated several times.

The example above merely illustrates an approach; other methods would have solved this problem more quickly. The advantages of simulation become outstanding if errors with different characteristics interact in complex patterns to influence the final result.

PRACTICAL APPLICATION

The advantages of simulation were exploited in the analysis of the following, forest inventory task. This project also showed how complicated the effects of errors become in even relatively simple operations.

THE INVENTORY TASK

An estimate of the total timber volume on a 1,060 square-mile area was to be obtained. This was achieved with little ground work, by relying on large-scale aerial photographs. Al-

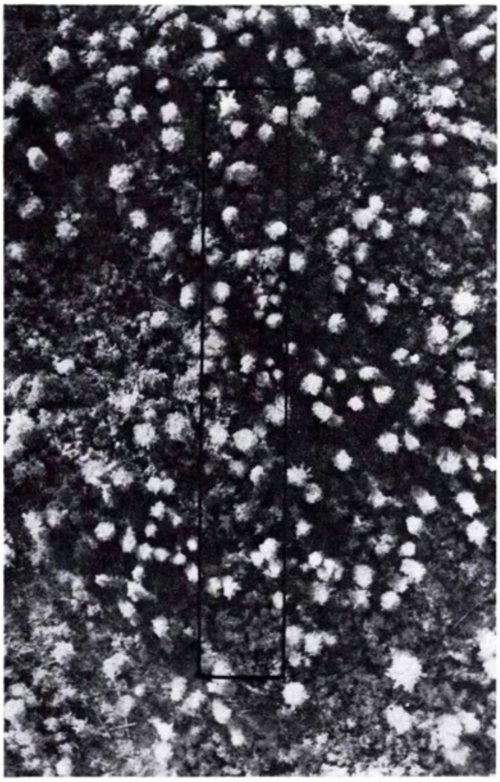

FIG. 1. A sample plot located on a large-scale photograph (see Sayn-Wittgenstein and Aldred, 1969, Fig. 4).

TABLE 2. SIMULATION OF INVENTORY PROCEDURE

1. Photographs taken, flying height observed for each photo.	$H^* = H \pm \Delta_H$
2. Parallax and differential parallax measured for each tree.	$P^* = P \pm \Delta_P$ $dp^* = dp \pm \Delta_{dp}$
3. Individual tree heights are calculated.	$h^* = H^* \times dp^*/P^* + dp^*$
4. Tree species identified.	Determine, for each tree, using probability matrix (Table 5) whether identification is correct or wrong.
5. Individual tree volumes calculated, using Equation 1: coefficients a, b differ with species.	Tree volume (v^*) estimated using Equation 1; if tree was wrongly identified in Step 4, above, use 1 with coefficients a, b for wrong species.
6. Some trees have not been seen by the interpreter.	Determine, for each tree, using h and the matrix of Table 6, whether the tree was missed; if it was missed, reduce its v^* to zero.
7. Plot volumes, V, calculated as sum of tree volumes.	$V^* = (\Sigma v^*)(H/H^*)^2$ The factor $(H)^2/(H^*)^2$ is included to simulate the effect of Δ_H on plot area (see text).
8. The true plot volumes V, and their mean, are compared with the estimated values V^* and their mean to yield an estimate of survey accuracy.	

Notation: H, P, dp, h, v, and V are respectively the true values for flying height, absolute parallax, differential parallax, individual tree height, individual tree volume, and plot volume. An asterisk * signifies the corresponding quantities after the effect of measurement or sampling errors; Δ_P, Δ_H, Δ_{dp} are the errors made in measuring P, H, and dp.

together, 126 randomly located sample locations were photographed at the very large scale of 1:2,000; at each of these locations a sample plot (Figure 1) was established. The height of each tree on these plots was determined photogrammetrically and its volume was then estimated from an equation of the form:

$$\text{Tree volume } (v) = \text{antilog } (ah + b) \qquad (1)$$

where h is photo-measured tree height and a and b are regression coefficients which vary with tree species and geographic location (Table 2). (More detailed descriptions of the forestry and sampling aspects of this task are given by Sayn-Wittgenstein and Aldred, 1969.) After estimates of individual tree-volumes were obtained they were summed to provide plot volume estimates, and finally estimates of timber on the entire area.

THE PROBLEM

An estimate of the accuracy of this inventory was required. If it could have been assumed that the individual plot volume estimates were *observations*, free from error; then a determination of survey precision would have been a relatively simple statistical task.

But many errors, particularly measurement errors, affect the individual plot estimates and the effect of these errors had to be known before the survey could be fully evaluated. Specifically $\overline{E}_V$, the root-mean-square error of plot volume, was required:

$$\overline{E}_V = \sqrt{\left\{\left[\sum_1^N (V_i^* - V_i)^2\right] \Big/ N\right\}} \qquad (2)$$

where N is the number of plots and V_i^*, V_i are, respectively, the estimated and the true timber volume on the ith plot.

The simulation model was thus primarily intended to yield $\overline{E}_V$, but also, in the course of more detailed analysis, to evaluate the impact of individual error sources.

THE SIMULATION MODEL

The data available were measurements of flying height, and absolute parallax and differential parallax of individual trees on the 126 plots used in the inventory. Respectively, these measurements were accepted as the true values of flying height H, parallax P and differential parallax dp. They were to play the same role in the simulation model as the variable B in Table 1. Also, at this stage all photo-identifications of species were accepted as correct, and it was assumed that no trees had been missed or erroneously included in plots.

The model of the real population was thus complete except for values of v, the true volumes of individual trees. These were not available and realistic values thus had to be artificially generated. However, using Equation 1, a regression estimate of the average volume $\hat{v}$ corresponding to each value of h was available. Values of v were generated by adding to each such estimate a deviate e drawn from a population with mean zero and a standard deviation equal to s_v, the known standard error of regression Equation 1,

$$v = \hat{v} + e.$$

Note that this provides, for each of the 1,400 trees involved, a realistic value of true volume. Thus, the stage is set for simulating the measurements of the various parameters given in the model.

The various stages of the inventory and the accompanying steps of the simulation process are summarized in Table 2. Of course, an almost endless number of sources of errors can influence the inventory result, and the simulation process will only describe the more important ones. The various stages of this process will now be followed through to deal with some of the more important sources of error.

DETERMINATION OF FLYING HEIGHT

Any error in flying height will have two principal consequences: (1) An error immediately occurs in the height of each individual tree because in the parallax formula of Stage 3, Table 2, H is replaced by H^*. The resulting error in tree height is directly proportional to Δ_H, the error in flying height. (2) In addition, error will occur in the area of sample plots laid out on photographs, because the photographic scale is assumed to equal f/H^* (where f is the focal length of lens used), whereas it actually equals f/H. For example, if $\Delta_H = +5\%\ H$, and an attempt is made to lay out on a photograph a square sample plot equivalent to 100 $\times$100 ft. on the ground, the actual area laid out will be 95.24$\times$95.24 ft.—an error of -9.30% in area. This error will be encountered in the simulation model again at Stage 7 of Table 2.

The mechanics of simulating an error in flying height are simple; an artificially generated *error* Δ_H is added to the parameter H, for each photograph to obtain the *measured* height H^*,

$$H^* = H + \Delta_H$$

and, just as in Table 1,

$$B + e = B^*.$$

To generate values of Δ_H it is important to know the characteristics of these errors. The simulation model could readily accommodate a multitutde of situations, including the following:

(a) Δ_H could follow a random, normal distribution about a mean of zero, in which instance realistic values for the standard deviation of the population of Δ_H would have to be established; (b) bias, possibly due to an error in calibration, might affect the reading of altitude—Δ_H would have to be generated from a population with a mean other than zero; (c) Δ_H could be a function of H, or of other parameters; (d) there could be other restrictions on Δ_H, for example, errors beyond a certain maximum might be impossible, negative errors might be impossible, or there could be a serial correlation between successive values of Δ_H.

ERRORS IN PARALLAX Δ_p AND DIFFERENTIAL PARALLAX Δ_{dp}

The treatment of these errors in the simulation process (Stage 2 of Table 2) is analogous to that for errors in flying height, described previously. It is first necessary to know the characteristics of the distribution of Δ_P and Δ_{dp}. As for Δ_H, a great variety of conditions can be satisfied by the computer in the process of generating errors, Δ_P and Δ_{dp} which are then added to the true values (P and dp) to obtain measurements.

In the absence of definite evidence to the contrary, Δ_P and Δ_{dp} have been assumed to be independent, with normal, random distribution about a mean of zero. This is a common assumption is photogrammetry, but certainly not an exact one. The simulation model would readily accommodate to evidence of more complicated characteristics.

The immediate effect of errors in P and dp will be on the estimation of tree height and then on all subsequent estimates of tree and stand volume.

ESTIMATION OF TREE HEIGHT

The calculation of tree height followed the parallax formula (Table 2); the measured tree height h^* was obtained from H^*, dp^*, and P^*, the values which included the measurement errors.

One interesting effect of random errors in parallax on the estimate of mean height was revealed: random errors introduce a slight bias. This is clearly shown by Table 3, which gives the result of a simulation run with 1,400 trees in which all errors, except errors in P were assumed to be zero; errors in P were assumed to be normally distributed, with mean zero and standard deviation σ_P. Note how the values of the average measured height, $\bar{h}^*$, increase with rising values for σ_p. That this should be true is readily illustrated:

assume a true value $P=2$ is subject to random errors $+1, -1, +1, -1$ to yield the readings 3, 1, 3, 1. The true value for $1/P=0.50$, the estimated mean value for P remains 2.0, but the mean of the reciprocals of the individual estimates is $(1/3+1+1/3+1)/4=.665$, well above 0.50. The effect of this phenomenon on the parallax formula leads to overestimates of h.

The effect of the interaction of errors in dp, P, and H on the estimated height of individual trees can also be readily demonstrated, as has been done in Table 4. Here can be seen the change in the mean-square error of individual tree height as the standard deviations of Δ_H, Δ_{dp} and Δ_P (σ_H, σ_{dp} and σ_P, respectively) assume different values. Again, the tabular values assume that the various errors are independent; it would be entirely feasible to simulate the effect of correlated errors.

ERRORS IN SPECIES IDENTIFICATION

The coefficients of the volume-estimating Equations 1 vary for different tree species. If a black spruce is misidentified as balsam fir the subsequent volume estimate will also be in error, even if all measurements are correct.

Guides to the nature of these species identification errors are known from past experience. Generally they are a function of tree species (some species are more difficult to recognize than others) and size (large, mature trees tend to have developed the species' characteristic crown form; young trees are less characteristic and often grow in dense clumps). For any combination of tree species and size (height) one can draw up a matrix, such as Table 5, to summarize the probabilities of identification errors. Spacing and crown canopy density are other factors affecting interpretation; they could be similarly treated by simulation.

It is not difficult to write a computer routine to simulate the interpretation of individual trees, using these probabilities. For example, Table 5 shows that a 20-foot tree of species A has a 75-percent probability of being correctly identified, and a 25-percent probability of being identified as species B. If such a tree is reached in the model, a pseudo-random number is generated from the range 1 to 100. If the number is 75 or less the tree has been correctly identified, if it is greater the tree is henceforth treated as if it had been identified as B.

There is nothing theoretically complicated about Table 5; in practice it represents information that because of its bulk is difficult to absorb in an error analysis by any method except simulation.

TABLE 3. EFFECT OF RANDOM ERRORS IN PARALLAX

σ_P (*mm*)	$\bar{h}^*$ *Estimated Mean Tree Height* (*m*)
0.0	14.14
0.5	14.14
1.0	14.16
2.0	14.40
3.0	14.65
4.0	15.23

TABLE 4. THE EFFECT OF ERRORS IN P, dp AND H ON THE ROOT-MEAN-SQUARE ERROR OF TREE HEIGHT, $E_{\bar{h}}$ IN METERS

σ_P (*mm*)	σ_{dp} (*mm*) 0.0	0.005	0.010	0.020
0.0	0.047	0.195	0.386	0.726
0.5	0.430	0.471	0.577	0.834
1.0	0.913	0.936	0.986	1.170
2.0	1.936	1.948	1.991	2.058

$\sigma_H=2.00$ meters; mean flying height $=650$ m.
$E_{\bar{h}}$ is analogous to $\overline{E}_V$ in Equation 2; h replaces V and $N=$number of trees.

ERROR DUE TO TREE VOLUME REGRESSION

Tree volumes used in the inventory are estimated with the help of Equations 1 derived by least-squares regressions of precise ground-measurements of volume on photo-measured heights of individual trees. In-

TABLE 5. ERRORS IN TREE SPECIES IDENTIFICATION

Actual Species	*Height Range* (*m*)	$P(A)$	$P(B)$	$P(C)$
	0–20	.75	.25	.00
A	21–40	.83	.17	.00
	41+	.95	.05	.00
	0–20	.20	.80	.00
B	21–40	.08	.92	.00
	41+	.01	.99	.00
	0–20	.08	.10	.82
C	21–40	.01	.09	.90
	41+	.00	.05	.95

$P(A)$, $P(B)$, $P(C)$ are, respectively, the probabilities of being identified as A, B, or C.

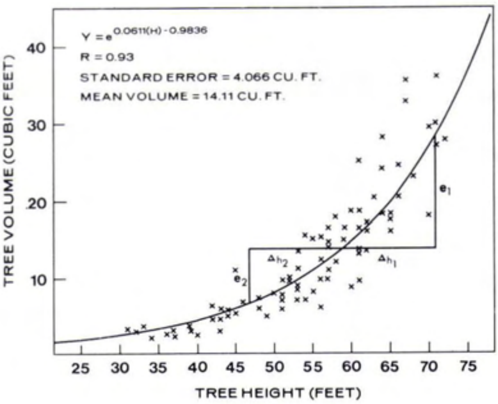

FIG. 2. Function used to estimate tree volumes (modified from Sayn-Wittgenstein and Aldred, 1969, Fig. 5).

dividual tree volume was calculated as a function of height: this function took the form represented in Figure 2. In many inventories other variables, such as crown area, would also have entered the equation.

Such regressions, of course, only show the average volume expected for any given height; the actual individual volumes tend to be dispersed about this line to a degree that is expressed by the regression's standard error s_v, a value calculated from the regression data and thus known in any inventory. Of course, the greater the value s_v has, the lower the precision is of any volume estimate. The inaccuracies introduced by s_v are covered in the simulation process: the volume estimate v^* used for an individual tree is the regression estimate corresponding to the measured height h^*. But, s_v has been used to derive the true volumes v (see above) and thus the magnitude of s_v will determine the difference $(v-v^*)$ which the simulation model can yield for each tree.

Another complication that arises from errors in estimating h is dealt with by simulation: because the function (Figure 2) is a sharply rising exponential, an overestimate of h (Δh_1 in Figure 2) will have a greater effect on the volume estimate than an equal underestimate Δh_2; note that $e_1 > e_2$. Thus even random errors in estimating h will lead to bias in volume estimation.

ERROR DUE TO MISSING TREES

In any sample using large-scale aerial photographs a certain number of trees will not be seen by the interpreter. From empirical evidence it is known that the probability of missing a tree is to a large extent, a function of its size; the information listed in Table 6 is thus obtained. Similar data are available to treat other factors, such as crown canopy density, which might affect the probability of missing a tree.

TABLE 6. PROBABILITY OF MISSING TREE ON PHOTOGRAPH

Tree Height (m)	*Probability of Being Missed*
0–5	0.12
6–10	0.06
10–20	0.02
20+	0.00

The probabilities listed can be incorporated in the simulation process just as the essentially similar matrix of Table 5. For example, Table 6 gives a 2-percent chance that a 19-meter tree will be missed. Thus, when the computer encountered a 19 m tree, it generated a pseudo-random number within the range 1 to 100. If the number was 1 or 2 the tree's volume was set equal to zero and the omission of the tree had been simulated.

EFFECT OF Δ_H ON PLOT AREA; CALCULATION OF PLOT VOLUME

As mentioned above, an error in flying height not only affects tree volume estimates through wrong estimates of tree height; it also leads to an error in plot area.

The effect of this error on the volume estimate can be approximated by assuming that the error in timber volume on the plot is directly proportional to the error in area. The following calculation is thus performed in the simulation program:

$$V^* = \left(\sum v^*\right)(H/H^*)^2 \qquad (3)$$

where Σv^* is the sum of the v^*'s for each plot. V^* is then the final estimated plot volume arrived at through simulation; it must be compared with V to estimate the survey's accuracy.

ESTIMATION OF ACCURACY

At the beginning of the simulation process we had set ourselves the limited goal of calculating the quantity $\overline{E}_V$ (Equation 2)—the mean-square error of the volume estimate. This can now be done because V^* and V, the estimated and the true volume, are available for every plot. Calculation of this error can be completed by considering all sources of error together or by isolating only a few.

A typical example of the form in which the

TABLE 7. ROOT-MEAN-SQUARE ERROR OF PLOT VOLUME ESTIMATES
$\overline{E}_V$ (cubic feet)

σ_P (*mm*)	σ_{dp} (*mm*) 0.0	0.005	0.010	0.015
0.00	27.8	27.3	26.9	28.7
0.05	28.1	27.7	27.3	29.1
1.00	65.4	65.9	68.4	69.0
2.00	367.6	377.5	385.7	394.1

$\sigma_H = 4.00$ meters.
$s_v = 4.07$ cubic feet, with the restriction that true tree volumes could not be less than 1.5 cu. ft.
Mean plot volume, $\overline{V} = 233$ cu. ft.

error summaries can appear is Table 7 which gives $\overline{E}_V$ for stated values of σ_P and σ_{dp} and $\sigma_H = 4.00$ m, $s_v = 4.07$ cu. ft.

Results, such as Table 7 are roughly equivalent to the results of a single experiment, and it is difficult to estimate their reliability. The simulation run may have to be repeated several times to judge the consistency of the results.

CONCLUSIONS

The purpose of the analysis was not to present the solution to a particular problem, as it is of limited general interest, but rather to illustrate how complex the error interactions tend to be and how computer simulation techniques can reveal their effects. Involved are some of the complex interactions and unexpected effects: an error in flying-height estimation affects tree height measurements and is confounded with an error in area estimation; *random* errors in parallax bias height estimates; *random* errors in tree height estimation bias volume estimates; errors in species identification and errors of omitting trees need to be included in the final calculation of accuracy of volume estimates; sampling errors combine with measurements errors to affect accuracy. All these sources of error can be dealt with in a simulation model.

The ability of simulation to deal with such situations is a great incentive to investigate the true nature of errors, because the presence of complex error characteristics will no longer be a handicap in studying error propagation. Realistic estimates of the errors that are incurred can be entered into the simulation process and an estimate of a survey's accuracy can be obtained after its completion. Of even greater importance is the possibility of increasing survey efficiency by establishing the effect that a change in any error would have on final accuracy.

For example, with reference to the inventory example: should effort be concentrated on improving the accuracy of species identification, or could more be gained by reducing s_v or Δ_H? This, of course, has always been able to be accomplished to some extent by analytical methods and experimentation. But with computer simulation it can be done more realistically and for much more complex problems.

ACKNOWLEDGEMENT

The authors acknowledge the valuable assistance of Messrs. H. M. Mucha and G. T. Maloley, Computer Programmers, Forest Management Institute.

REFERENCES

Aldred, A. H. 1967. Diagnosis of errors affecting the accuracy of tree-volume estimates from large-scale air photos. *Proceedings—2nd Seminar on Air Photo Interpretation in the Development of Canada*, Ottawa, March 13–15, 1967, pp. 122–130.

Gerrard, D. J. 1969. Error propagation in estimating tree size. *Photogrammetric Engineering*, Vol. 35, No. 4, pp. 355–362.

Sayn-Wittgenstein, L. and A. H. Aldred. 1969. A forest inventory by large-scale aerial photography. *Pulp and Paper Magazine of Canada*, Vol. 70, No. 17, pp. 92–95.

Schut, G. H. and M. C. van Wijk. 1965. The determination of tree heights from parallax measurements. *The Canadian Surveyor*, Vol. 19, No. 5, pp. 415–427.

Whiteside, A. E. and D. B. Lipski. 1968. Computer simulation of automatic stereoplotters. *Bendix Technical Journal*, Summer 1968, pp. 53–58.

Wolters, J.A.M. 1968. Accuracy of analytics by computer simulation. *Photogrammetric Engineering*, Vol. 34, No. 10, pp. 1084–1092.

Yates, F. 1960. *Sampling methods for censuses and surveys*. 3rd edition. (Charles Griffin, London; Hafner, New York), 440 p.

ASP Needs Old Magazines

Because of an unexpected demand for journals and student requests, the supply of some back issues of PHOTOGRAMMETRIC ENGINEERING has been depleted. Consequently, until further notice, National Headquarters will pay to the Regions—or to individual members—$1.00 for each usable copy of the following issues sent to Headquarters, 105 N. Virginia Ave., Falls Church, Va. 22046:

1969 January, March, June and July