Analytica Chimica Acta, 157 (1984) 273—290
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

A SOFTWARE PACKAGE FOR THE GENERATION OF NOISE WITH WIDELY DIVERGENT SPECTRAL PROPERTIES
The Simulation of Realistic Stationary Detector Noise

J. M. LAEVEN and H. C. SMIT*

Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam (The Netherlands)

J. V. LANKELMA

Stichting Mathematisch Centrum, Department of Applied Mathematics, Kruislaan 413, 1098 SJ Amsterdam (The Netherlands)

(Received 2nd August 1983)

SUMMARY

The computer program discussed is capable of generating stationary noise with optional statistical parameters. Five standard noise types can be generated: white, 1/f, first order, Gaussian, and damped cosine noise. The input to the program can be a measured or arbitrarily chosen autocovariance function (or power spectrum). The autocovariance function (power spectrum) computed from the generated noise will be in perfect agreement with the input autocovariance function, if an infinite number of noise data is generated. Some theoretical work on noise generation on the basis of the autocovariance function is described. The autocovariance function of 1/f noise is theoretically derived from the power spectrum, based on a model described earlier. A few examples of the use of the program and an example of a possible application are given. A peak-finding procedure is tested with a simulated chromatogram contaminated with different types of noise. Applications are possible in data processing, information extraction, simulation and automation.

The use of computers in data analysis has created possibilities of developing information extraction methods (i.e., extraction of relevant information from analytical signals) in a systematic manner. Consequently, determining and reducing the influence of noise is gaining importance. However, the relative complexity of the mathematics involved in data analysis to yield optimal results has been a limiting factor. The development of chemometrics can be partly explained from this background. To date, information extraction methods in analytical chemistry were usually based on, and tested (if at all) with, a white or first-order noise model. It is known, however, that many detectors exhibit noise characteristics different from the white or first-order model [1]. This means that the algorithm originating from this approach will be far from optimal in practice and even may give erroneous results.

The fields in which reliable noise models are necessary are very diverse. In data processing, filter procedures have to be developed and tested with use

0003-2670/84/$03.00

of realistic noise models to yield optimal performance. In information extraction methods, these models are necessary for qualitative analysis (e.g., peak-finding procedures in chromatography; interpretation of spectra) and quantitative analysis (e.g., determination of peak area and retention times in chromatography). Further, in simulation, when processes are simulated using systems theory, the need of reliable noise models is obvious. Also in the field of automation, noise models play an important role; programs to interpret spectra automatically have to be developed with use of different noise types to minimize "misses" and "false alarms".

In this paper a program capable of generating stationary noise with selectable and realistic characteristics is presented. This program can be of much help in all the research fields mentioned above.

THEORETICAL BACKGROUND

A set of independent numbers u_j, representing white noise with statistical properties, is defined as follows

$$E[u_j] = 0 \tag{1}$$

$$E[u_i u_j] \equiv R_{uu}(i-j) = \delta_{i-j} \tag{2}$$

where $E[u_j]$ represents the expected value of u_j (mean value), δ_{i-j} is the delta Dirac distribution, and $R_{uu}(i - j)$ is the autocovariance function (ACVF) of u. (Symbols are defined in Table 1.) The general definition of the ACVF of a stationary stochastic variable y is given by

$$R_{yy}(\tau) \equiv \lim_{T\to\infty} (1/T) \int_{-T/2}^{T/2} y_t y_{t+\tau} \mathrm{d}t \tag{3}$$

A (discrete) system with impulse response h_k, where k represents (discrete) time, is also defined. This means that if a Dirac pulse δ is applied to the input of the system, then the output of the system equals the impulse response h. It is well known from signal theory that any signal can be represented by a series of Dirac pulses [2]. If such a (discrete) signal is applied at the input of the described system, then the output of the system can be found by summing the responses on each distinct pulse. If this procedure is followed for white noise at the input of the system, the result at the output will be the convolution summation

$$x_l = \sum_{k=-\infty}^{\infty} h_k u_{l-k} = \sum_{k=1-M}^{M-1} h_k u_{l-k} \tag{4}$$

In practice, the impulse response of a system is finite in time, thus infinite summation can be replaced by finite summation over $2M-1$ values, where M is sufficiently large, as is done in Eqn. 4. The convolution means that the responses of the system to all input values in the past are summed to obtain the output of the system.

TABLE 1

List of symbols and definitions

u_j	White noise number
$E[\ ]$	Expected value of the expression within the brackets
$R_{xx}(\tau)$	Autocovariance function of the function x_t with time lag τ
$R_{uu}(i-j)$	Discrete autocovariance function of the (discrete) stochastic variable u
δ_i	Delta Dirac distribution, for the discrete case: $\delta_i = \begin{cases} 1 & i = 0 \\ 0 & i \neq 0 \end{cases}$
x_l	Discrete filter output
h_k	Filter impulse response
$2M-1$	Number of significant impulse response coefficients
FT	Fourier transform
FT^{-1}	Inverse Fourier transform
$H(\omega_l)$	Fourier-transformed impulse response
$G(\omega)$	Spectral power density function
$E_1(\xi)$	Exponential integral
t_0	Expected time duration of pulses making up the current in a flame ionization detector
K	Detection threshold parameter
ϕ_t	Function describing an asymmetric peak
H	Peak top amplitude
A	Asymmetry factor
t_R	Retention time
σ_p	Gaussian standard deviation of a peak
σ_n^2	Variance of baseline noise
S/N	Signal-to-noise ratio, defined as amplitude max/σ_n
$\bar{x}$	Mean value of a random sample of the distribution x
μ	Mean value of the distribution x
$s_{\bar{x}}$	Standard deviation of the mean value of a random sample of distribution x
T	Observation time
T_x	Time constant

The statistical properties of the system output (given by Eqn. 4) can be evaluated as follows

$$E[x_l] = \sum_{k=1-M}^{M-1} h_k E[u_{l-k}] \overset{(1)}{=} 0 \tag{5}$$

$$R_{xx}(i) = E[x_l x_{l+i}] \overset{(4)}{=} \sum_{j=1-M}^{M-1} \sum_{k=1-M}^{M-1} h_j h_k E[u_{l-k}\, u_{l+i-j}]$$

$$\overset{(2)}{=} \sum_{j=1-M}^{M-1} \sum_{k=1-M}^{M-1} h_j h_k \delta_{l-j+k} = \sum_{k=1-M}^{M-1} h_k h_{k+i} \tag{6}$$

(Numbers placed on the equal signs refer to eqns. derived before)

The last part of Eqn. 6 is known as the (discrete) system covariance function. Thus, the result is that the ACVF of the signal x_l (the output of the system) is equal to the system covariance function.

Equation 6 can be used to compute a shaping filter with impulse response h, which filters white noise in such a manner that the ACVF of signal x at the output of the shaping filter is in the desired form. So, if noise with a

certain ACVF is to be generated, the impulse response h of the shaping filter must be computed using Eqn. 6. This impulse response can be obtained by Fourier transformation of Eqn. 6

$$\mathrm{FT}[R_{xx}(i)] = \mathrm{FT}\left[\sum_{k=1-M}^{M-1} h_k h_{k+i}\right] \tag{7}$$

Because $R_{xx}(i)$ is even in i, Eqn. 7 will hold if i is replaced by $-i$. Thus the convolution theorem, i.e., convolution in the time domain is multiplication in the frequency domain, can be applied to Eqn. 7 and results in

$$\mathrm{FT}[R_{xx}(i)] = H^2(\omega_l) \tag{8}$$

where $H(\omega_l)$, the complex frequency response, represents the Fourier-transformed impulse response h_k. From this expression the impulse response of the shaping filter can be calculated

$$H(\omega_l) = \{\mathrm{FT}[R_{xx}(i)]\}^{1/2} \tag{9}$$

$$h_k = \mathrm{FT}^{-1}\{\mathrm{FT}[R_{xx}(i)]\}^{1/2} \tag{10}$$

Thus, summarizing, the impulse response of the shaping filter, generating noise with the required ACVF properties, can be found by Fourier transformation of the ACVF, taking the square root, and Fourier back-transformation. Because of the symmetry of the functions involved, a fast Fourier cosine transformation will do. After evaluation of the impulse response in the described manner, the signal x_l with the desired ACVF characteristics can be generated according to Eqn. 4.

Flicker noise

Some detectors exhibit noise characteristics that may hamper the signal analysis and interpretation to a high extent. Flicker or 1/f noise, caused by electrical disturbances of an incompletely understood nature, is an example of a noise difficult to handle. The problem of this type of noise is that the power of the noise is inversely proportional to the frequency down to very low frequencies. This implies that the highest noise power occurs in the same frequency area where generally analytical signals exhibit their maximum power. An example of a detector with this noise characteristic is the flame ionization detector (FID) [1].

A realistic approximation model of the 1/f noise power spectrum, in which the (actually non-existent) singularity at $\omega = 0$ is avoided, has been given by Smit and Walg [3]

$$G(\omega) = c \int_{t_m}^{t_n} [1/(1 + \omega^2 t_0^2)]\, \mathrm{d}t_0 = c\, \{[\arctan(\omega t_n) - \arctan(\omega t_m)]/\omega\} \tag{11}$$

where t_n and t_m are the time constants, $t_n > t_m > 0$, t_0 is the expected value of time duration of the pulses making up the current in a FID, and c is some constant.

Fourier back-transformation of the 1/f power spectrum to obtain the autocovariance function of 1/f noise can be done with a cosine transformation, because $G(\omega)$ is even

$$R_{xx}(\tau) = \mathrm{FT}^{-1}[G(\omega)] = 2c \int_0^{\infty} \int_{t_m}^{t_n} [\cos(\omega\tau)/(1 + \omega^2 t_0^2)]\, \mathrm{d}\omega\, \mathrm{d}t_0 \tag{12}$$

Introducing a new variable $z = \omega t_0$ with $\mathrm{d}z = t_0\, \mathrm{d}\omega$, and using the identity

$$(2/\pi) \int_0^{\infty} [\cos(ax)/(1 + x^2)]\, \mathrm{d}x = \mathrm{e}^{-a}$$

rearrangement of Eqn. 12 yields

$$R_{xx}(\tau) = 2c \int_{t_m}^{t_n} \frac{1}{t_0} \int_0^{\infty} [(\cos z\tau/t_0)/(1 + z^2)]\, \mathrm{d}z\mathrm{d}t_0$$

$$= c\pi \int_{t_m}^{t_n} \exp(-\tau/t_0)\, \mathrm{d}t_0/t_0 = c\pi\, [E_1(\tau/t_n) - E_1(\tau/t_m)] \tag{13}$$

where E_1 is an exponential integral, the general definition of which is given by Abramowitz [4]

$$E_1(\xi) = \int_{\xi}^{\infty} (e^{-s}/s)\, \mathrm{d}s$$

The constant $c\pi$ in Eqn. 13 can be rewritten to obtain the ACVF in the appropriate form

$$R_{xx}(\tau) = \sigma^2\, [E_1(\tau/t_n) - E_1(\tau/t_m)]/\ln(t_n/t_m) \tag{14}$$

Numerically, the obtained ACVF can be evaluated by using polynomial and rational approximations given by Abramowitz [4].

Theoretical approximations of the power spectrum and the ACVF of 1/f noise are available now, so that it becomes possible to compare the characteristics of measured noise sources with those of 1/f noise in both the Fourier domain and the time domain. Furthermore, the ACVF of measured noise can be used to classify the noise type as 1/f without the necessity of Fourier transformation, where problems arise because a reliable power spectrum can mostly only be found if optimal windows are used.

As the ACVF and power spectrum characteristics of 1/f noise are now known approximately, it is useful to compare 1/f noise with first-order noise. If the two power spectra are compared, one can see that if the dominating time constants t_n and T_x (see Eqns. 14 and 15) are approximately equal, the power of 1/f noise in the very low frequency area and the high frequency area is larger than the power of first-order noise in those regions, whereas for the middle region the reverse is true.

DESCRIPTION OF THE SAMSON PROGRAM

The purpose of the FORTRAN program SAMSON (Spectrum/ACVF Matching Synthesis Of Noise) is to generate zero mean, stationary noise based on a predefined ACVF/power spectrum given as input to the program; this ACVF/power spectrum can originate from measured noise (e.g., detector noise), computed externally and subsequently read by SAMSON, or can be a standard ACVF generated by the program. Four common standard types of ACVF are possible

(1) first-order noise $\sigma^2 \exp(-|\tau|/T_x)$ (15)

(2) Gaussian noise $\sigma^2 \exp(-\tau^2/T_x^2)$ (16)

(3) damped cosine first-order noise $\sigma^2 \exp(-|\tau|/T_x) \cos \omega_0 \tau$ (17)

(4) "1/f noise" $\sigma^2 [E_1(\tau/t_n) - E_1(\tau/t_m)]/\ln(t_n/t_m)$ (18)

First-order noise can originate from shaping filters (or processes) described by a first-order differential equation (low-pass filters). Many amplifiers can be considered as first-order systems. Gaussian filters cause the second type of noise. The third type originates from a bandpass filter as used in lock-in amplifiers. Generally it can originate from filters described by negative discriminant second-order differential equations with very high $\omega_0 T_x$, allowing the sine term to be ignored in the solution. The fourth type, 1/f noise, is observed in many detectors and other noise sources. In all cases, the described types of noise are observed as the output of the shaping filter, if white noise is applied as input.

To extend the use of standard ACVFs, an option to change individual ACVF values is inserted in the program, e.g., to allow study of the noise characteristics if spikes periodically occur in the ACVF, as may be observed with detectors that are zero-adjusted during a series of measurements. Of course, this option is only useful if the desired ACVF does not deviate too much from one of the standard types. If the deviation is too large, the ACVF can better be generated externally, and subsequently read by SAMSON.

Summarizing, any ACVF/power spectrum can be used as input to SAMSON. The program will generate noise with properties according to this ACVF/power spectrum and store the noise on a disc file. The ACVF/power spectrum, used as input, and the output ACVF/power spectrum, calculated from the generated noise, can optionally be stored on disc too, permitting their use for other purposes. Furthermore, the input and output ACVFs/power spectra are plotted simultaneously in one graph, allowing careful inspection and comparison. Of course, the similarity of input and output ACVFs/power spectra is a statistical matter, thus depending on the number of noise data generated. In the computer system used (HP 1000, model 45-F, operating under RTE IVB) the configuration permits 16 384 noise data to be generated at one time as a maximum. The exact number depends on the number of weighting coefficients necessary to describe the filter,

because these coefficients and the noise data are stored in the same array for efficiency reasons. Because of the boundary conditions, imposed by the structure of the computer system, SAMSON is written as an independent program communicating with the user interactively. The number of possible options and the need for visual inspection and interpretation of the chosen input ACVF/power spectrum and the resulting noise justifies this choice. Furthermore, the interactive mode allows use of SAMSON for educational purposes.

In Fig. 1 a flow chart is given of the main parts of the program (written in FORTRAN IV). The required memory is 124 kbyte. SAMSON is built up of five main parts

(1) Main program: I/O control, overlay calling and library routines (fast Fourier transform, white noise generation); ACVF/spectrum computation; computation of the shaping filter;

(2) Overlay FNCGN: generation of standard ACVFs;

(3) Overlay DSCIO: reading of input ACVF/spectrum from file; filtering of the white noise using the computed shaping filter; storage of the generated noise on disc;

(4) Overlay DSCO2: storage input/output ACVFs/power spectra on disc;

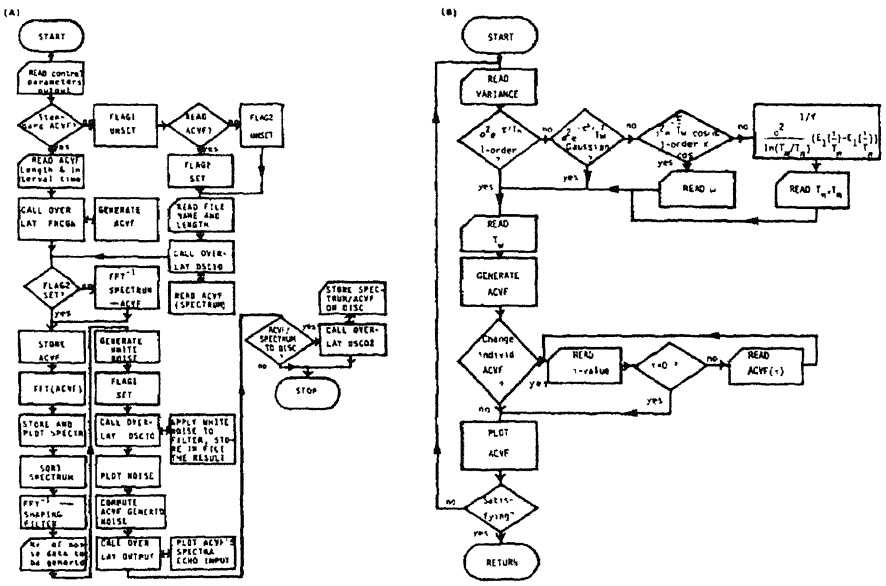

Fig. 1. (A) Flow chart of the main program of SAMSON; FFT stands for Fast Fourier Transform. (B) Flow chart of overlay FNCGN; four standard ACVF types are available for noise generation.

(5) Overlay OUTPT: Supply of plot facilities; output control (hard copy of input parameters, filter coefficients, used filter names etc.).

The time required for the noise generation should be considered. It can be seen in Eqn. 4 that each noise number has to be generated from $2M-1$ white noise numbers, where M depends on the ACVF length and on the number of significant filter coefficients (of course, these two criteria are related). For simplicity, it can be assumed that the filter is sufficiently described by 100 (M = 100) coefficients and that 10 000 noise data have to be generated. It follows that $\sim 2 \times 10^6$ multiplications/additions are needed. With conventional hardware/software this could result in rather time-consuming computations. However, in the computer system used, a vector instruction set is available, allowing considerable reduction in computation time. In the example, the total computation time is only 135 s. If RTE-VI is available, this time can be reduced further. In fact, RTE-VI permits some useful adaptation of the program; these are not described in detail here because their effect is rather system-dependent. The main adaptation is the extension of the number of data that can be generated up to 200 000 at the cost of some speed.

Confidence interval

To determine whether the ACVF of the noise generated by SAMSON fits the model used as input ACVF, the confidence intervals for the input ACVF and for the given number of noise data generated have to be estimated. This can be done using the Bartlett formula [5]

$$\sigma^2 [R_{xx}(\tau)] = [1/(T-\tau)^2] \int_{-T+\tau}^{T-\tau} (T-|r|-\tau) \{R_{xx}^2(r) + R_{xx}(r-\tau)R_{xx}(r+\tau)\} \, dr \qquad (19)$$

where r is an auxiliary variable (time).

In SAMSON optionally the confidence intervals (99.72%) are depicted: the ACVF of the generated noise should be within the interval $R_{xx}(\tau) \pm 3\sigma [R_{xx}(\tau)]$, where $R_{xx}(\tau)$ is the input ACVF.

RESULTS

In order to illustrate the range of possibilities of SAMSON, noise of all the mentioned standard types was generated. In addition, the ACVF of measured noise of different origins was used as input to the program, and the noise according to this ACVF was generated. The results are shown in Fig. 2. It can be seen that the agreement of the input ACVF with the ACVF of the generated noise is good. This is also true for the power spectra. The deviations are all within the Bartlett confidence intervals (Eqn. 19), which are not depicted in Fig. 2 to avoid confusion. In addition, for different numbers of noise data of the four standard noise types, the ACVFs obtained were never

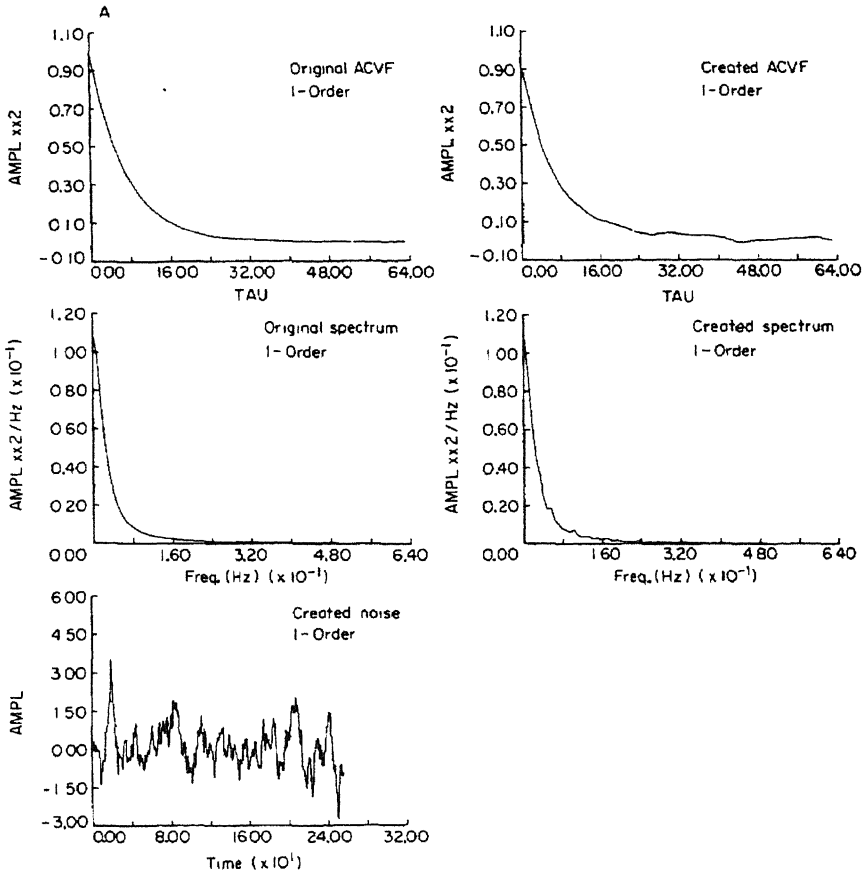

Fig. 2. Results of noise generation by SAMSON. At the left of each part (A—F) are shown the desired ACVF and the corresponding power spectrum, which are used as input to SAMSON; for ACVFs of measured noise, a record of the noise is also shown. At the right of each part, the resulting noise, its ACVF and power spectrum are depicted. Thus the similarity of input and output functions can be compared easily. (A) First-order noise (expression 15). (B) Gaussian noise (16). (C) Damped cosine noise (17). (D) 1/f noise (18). (E) Noise from the light signal leaving the monochromator of an inductively-coupled plasma emission spectrometer. (F) Noise originating from a flame ionization detector.

significantly different from the input ACVF used. The figures of the ACVFs and power spectra are based on 5000 noise data generated. Only a part of these 5000 numbers is shown in the figures. Table 2 lists the parameters for the standard ACVFs of Fig. 2.

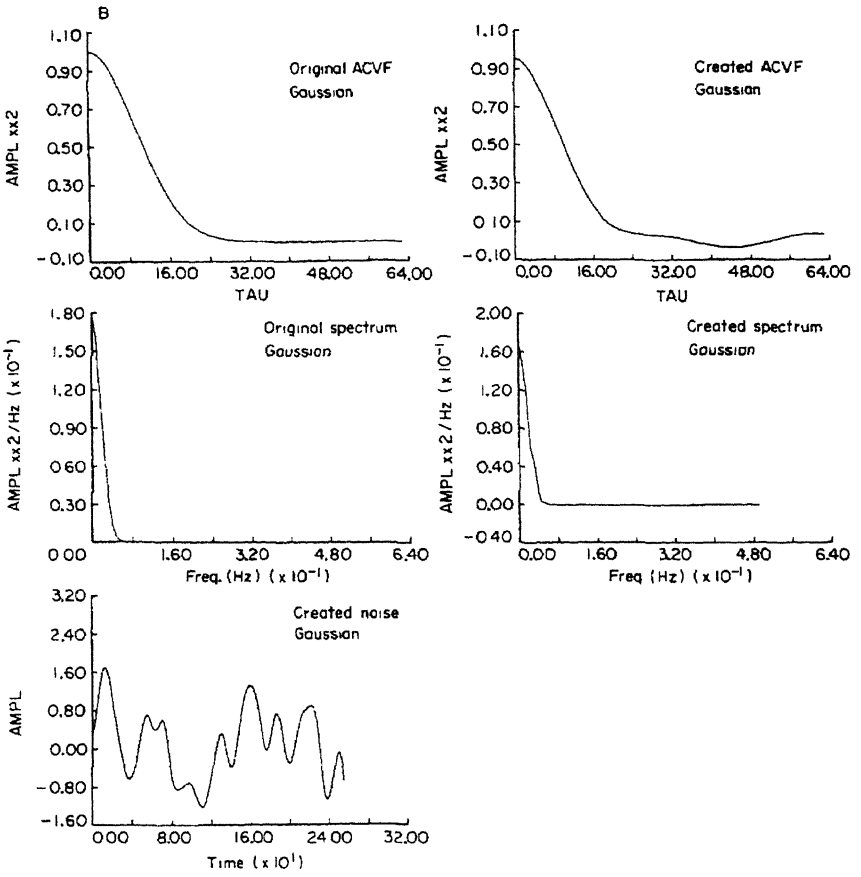

Fig. 2B.

APPLICATION TO TESTING OF PEAK-FINDING PROCEDURES

To demonstrate the possible applications of SAMSON, a realistic example is used. In chromatography, various attempts are being made to replace the usual visual interpretation of the chromatogram by automatic interpretation. Of course, to do so, the algorithm designer must formulate the problem of rather complex interpretation in a systematic manner. Many steadily improving peak-finding procedures have been developed. The main drawback of even the most sophisticated of these procedures is that the assumed noise models are generally not compatible with the noise observed in practice. For this reason, it is necessary to test the performance of the peak-finding procedure in use with different types of noise. SAMSON can readily be used for this purpose, and as a demonstration a first screening test of a peak-finding procedure is given below.

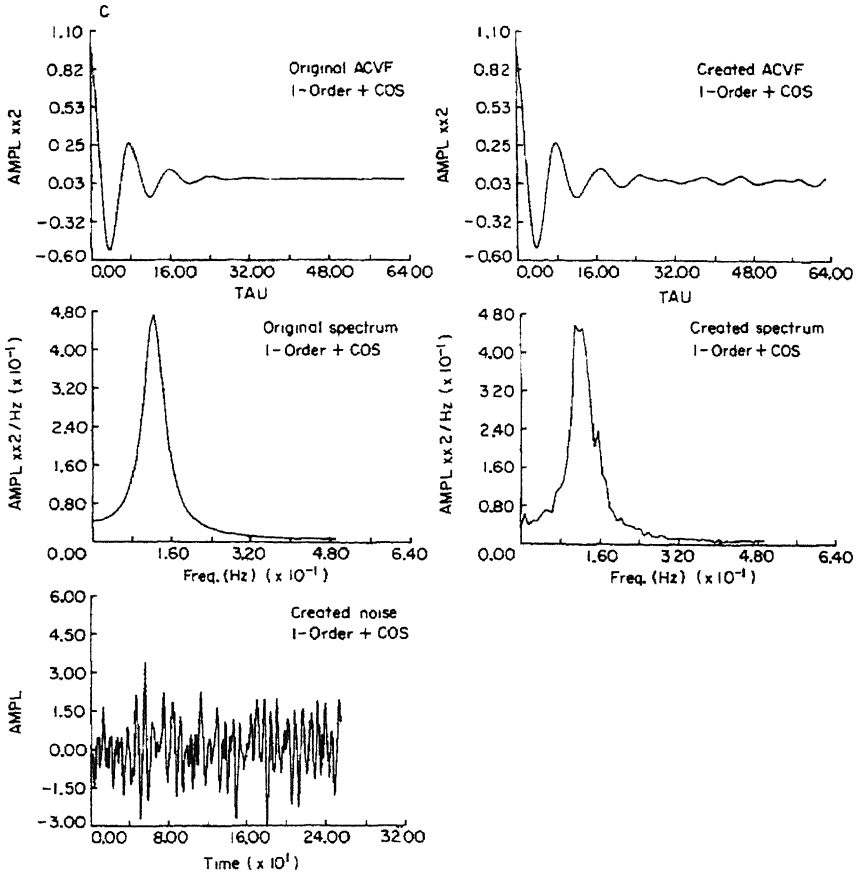

Fig. 2C.

A slightly modified version of the simple peak-finding procedure developed by Bobba and Donaghey [6] was used; the procedure was originally intended to handle chromatographic data with a microcomputer. The purpose of this test procedure is not to present an exhaustive statistical study on peak-finding procedure testing, but mainly to demonstrate the use of SAMSON.

Algorithm

The peak-finding program starts reading a certain preset number of data points, and evaluates the standard deviation and the approximated derivative ($\Delta x/\Delta t$) to determine linear drift. If these values do not exceed the preset values, the chromatographic system is considered to be stable. If they exceed the preset values, a new series of data is read and the procedure is

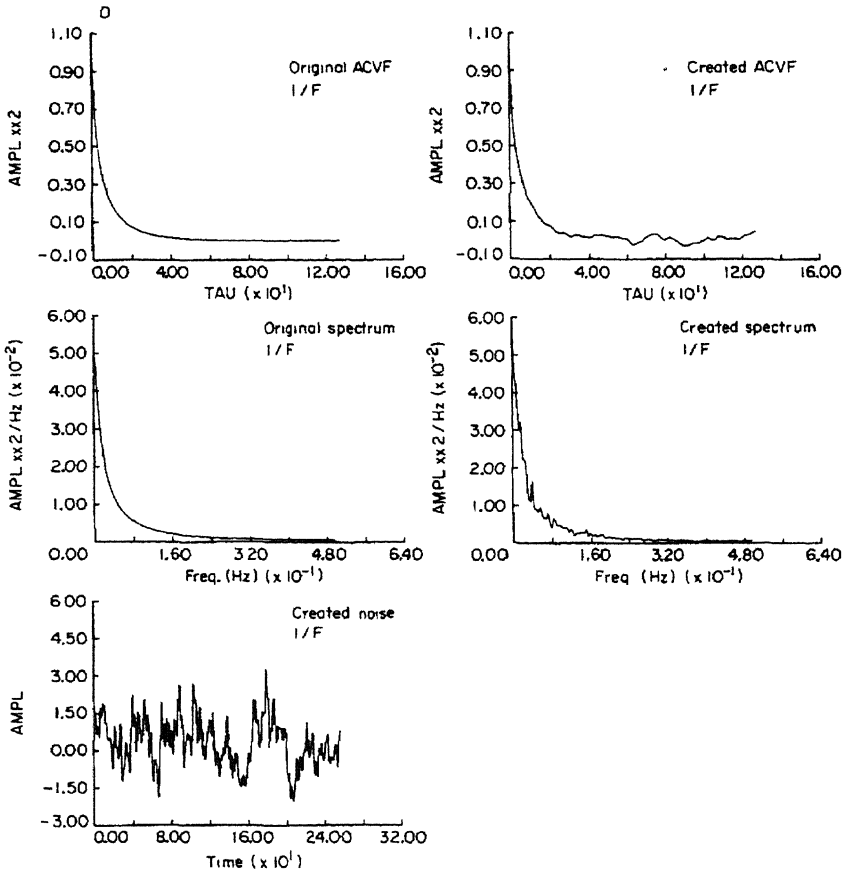

Fig. 2D.

repeated. After stabilization, detection of the first peak-start begins. A new preset number of data (equal to the preset parameter window length) is read and the standard deviation (σ_n) is evaluated. Peak start is detected if the last read value exceeds $K\sigma_n$, where K is a threshold parameter which can be chosen freely by the user (e.g., a read value > baseline + $3\sigma_n$ means that the probability a peak is present is >99.72%). If the last value does not exceed the threshold, another value is read until it does. After peak start detection, new values are read; the peak maximum is detected if a data point is read, which is $3\sigma_n$ smaller than the preceding value, which is considered to be the peak maximum. Peak end is detected if the approximated derivative of two successive values exceeds the slope determined at the peak start. The approximated derivative is defined as $x_j - x_{j-1} + K\sigma_n$. The peak-finding procedure is then repeated to the end of the data.

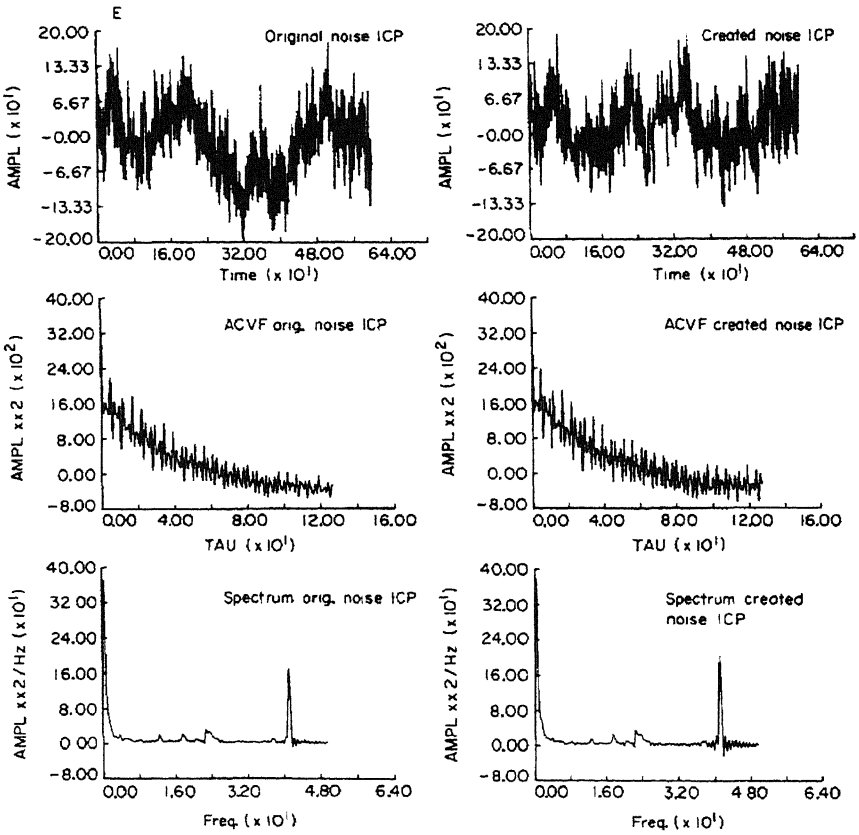

Fig. 2E.

Test conditions

A test chromatogram (400 data points) with 6 peaks was simulated (Fig. 3; Table 3) according to the model given by Fraser and Suzuki [7]

$$\phi(t) = H \exp\left([-\ln 2/A^2]\ \{\ln\ [1 + A\ (t - t_R)/\{\sigma_p\ (2 \ln 2)^{1/2}\}]\}^2\right) \qquad (20)$$

Peak starts, peak ends and the retention times were stored to be compared with the parameters detected by the peak-finding procedure. Peak end and peak start for the unresolved peaks were located at the minimum between the respective peak maxima.

Subsequently, five noise files of different types, containing 40 000 data each, were generated with SAMSON on the basis of an ACVF, length 64, and $\sigma_n^2 = 1.0$ (Table 4). The following test procedure was then chosen: for each of the five noise types, 400 data were added to the test chromatogram:

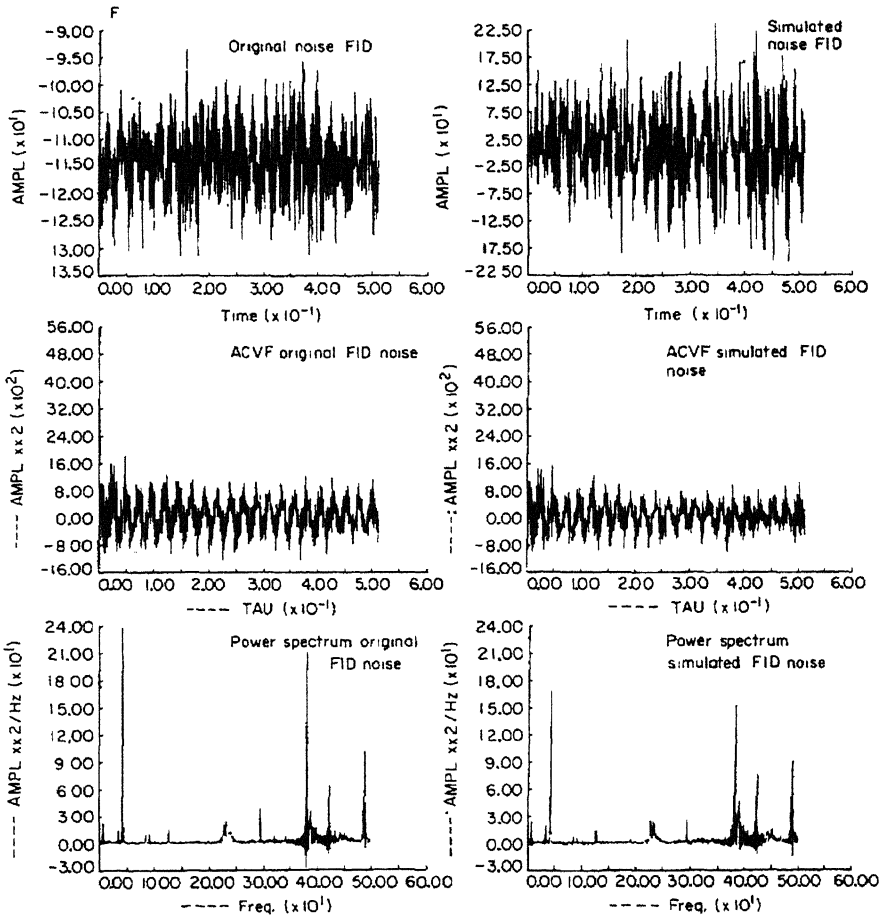

Fig. 2F.

noisy chromatogram = chromatogram + $p \times$ noise, where p is 1.0, 0.3 or 0.05. Thus the S/N ratios (maximum amplitude/σ_n) for the highest peak were about 10, 33 and 200. This test procedure was repeated 100 times with new noise data. Then, each noisy chromatogram was subjected to the peak-finding procedure six times, with input parameters: window length (4, 8 and 12), and K factor (threshold parameter, 1.0 and 2.0). If K was 3.0, bad results were always obtained in a preliminary study. The performance of the peak-finding procedure was quantified by awarding scores 1 to 10 for each peak parameter (peak start, peak maximum and peak end) correctly detected when a peak was in fact present. If the absolute difference ΔQ (time units) between the detected peak parameter and the real value was <10, then the detection was awarded $10-\Delta Q$, yielding a maximum score of 180 (six peaks, three parameters each), which was expressed as 100% performance.

TABLE 2

Parameters of the ACVFs shown in Fig. 2

Noise type	σ^2	T_x	ω_0
First order	1.0	8.0	
Damped cosine	1.0	6.0	0.77
Gaussian	1.0	12.0	
1/f	1.0	30 and 0.1	

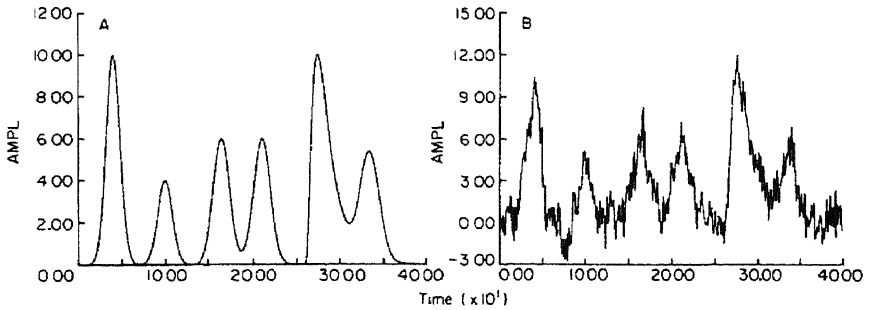

Fig. 3. (A) Simulated chromatogram to test the peak-finding procedure. (B) Test chromatogram with noise added (S/N for the highest peak is 10).

TABLE 3

Peak parameters of the chromatogram in Fig. 3

Peak No.	t_R	σ_{Peak}	A	Ampl. max.
1	40	8.0	0.0	10
2	100	8.5	0.0	4.0
3	165	9.0	0.0	6.0
4	210	9.5	0.0	6.0
5	275	10.0	0.7	10.0
6	335	10.5	0.0	5.0

Summarizing, the peak-finding procedure performance was tested with a chromatogram contaminated with five different noise types at three S/N levels. Further, the peak-finding procedure parameters, window length and threshold value, were varied. For each noise type, at each level, and for each individual K factor and window length, the test was repeated 100 times to obtain reliable results. Thus, altogether the peak-finding procedure was run 9000 times.

TABLE 4

Parameters of the ACVFs of the noise types used in the test chromatogram

Type of noise	σ_N	Time const.	ω_0	Spikes	
				τ	ACVF(τ)
First-order	1.0	8.0	—	—	—
Damped cosine first-order	1.0	8.0	0.555	—	—
Gaussian	1.0	3.0	—	—	—
1/f	1.0	0.1			
		10.0	—	—	—
First order + spikes	1.0	8.0	—	10	0.5
				20	0.2
				30	0.07
				40	0.02

Results

The scores of the peak-finding procedure under the different test conditions are summarized in Table 5. The mean score $\bar{x}$ (%) of the 100 runs is given. The 95% confidence interval on μ is $\bar{x} - 1.98\, s_{\bar{x}} < \mu < \bar{x} + 1.98\, s_{\bar{x}}$, where $0.4 < s_{\bar{x}} < 1.4$. The mean scores are significantly different at the 90% confidence interval if they differ (between noise types) more than 1.5, according to the Smith—Satterthwaite method [8].

A critical look at the table reveals some features of the peak-finding procedure under test. The deviation of the detected peak parameters averaged over the three S/N levels for K factor and window length where the procedure has optimal performance, is 5.5 (time units), which is about 0.6 σ_p. This implies a rather low detection performance. The deviation is almost independent of the S/N ratio: 6.1, 5.6 and 4.8 for S/N ratios of 10, 33 and 200, respectively.

TABLE 5

Performance of the peak-finding procedure applied to the test chromatogram contaminated with noise of different types. The score of the procedure is 100 if all peak parameters are detected without errors

σ_{Noise}	1.0						0.3						0.05					
K factor	1			2			1			2			1			2		
Window	4	8	12	4	8	12	4	8	12	4	8	12	4	8	12	4	8	12
First order	21	35	35	29	36	36	24	35	39	34	42	41	23	36	43	35	48	51
Spiked first order	19	35	37	23	38	39	21	37	42	28	43	43	18	37	43	32	50	51
Gaussian	25	34	39	32	34	33	27	36	38	34	39	38	23	36	39	39	51	49
Damped cosine	39	44	44	37	29	32	35	46	48	39	41	42	29	47	50	44	53	53
1/f	21	39	37	31	35	35	23	45	45	38	46	44	24	42	47	37	52	52

If the performance of the peak-finding procedure is considered in relation to the type of noise, some different features can be observed. First, it can be seen that the performance is approximately equal if the chromatogram is contaminated with first-order or spiked first-order noise; the optimal procedure parameters, K factor and window length, are about equal in both cases. The same is true for a chromatogram contaminated with Gaussian or 1/f noise, although the performance with 1/f noise is better. Table 6 gives an overview of optimal parameter combinations for the different types of noise.

The test procedure described shows some features of the peak-finding procedure under different noise circumstances. The procedure yields the parameters that have to be chosen to obtain the best performance for each type of noise and S/N ratio. An overall conclusion is that the algorithm for the peak-finding procedure needs improvement, because its performance is poor. Of course, simple addition of noise to simulated chromatograms is no guarantee of a realistic simulation of real chromatograms: care has to be taken that the peak profiles are not affected significantly by the filter. Thus, the chosen noise characteristics should be in accordance with the profile of the chromatogram.

Of course, the test procedure is too simple to obtain an exact picture of the behaviour of the peak-finding procedure; for example, "misses" and "false alarms" are not detected, nor is the performance as a function of peak resolution. Yet, the test procedure appears to be satisfactory as a preliminary screening method.

TABLE 6

Recommended window length and threshold value for different types of noise and S/N ratios

Type of noise	S/N	K-factor		Window	
		1	2	8	12
First-order	10		x	x or x	
	33		x	x or x	
	200		x		x
Spiked	10		x	x or x	
first-order	33		x	x or x	
	200		x	x or x	
Gaussian	10	x			x
	33		x	x or x	
	200		x	x or x	
Damped cosine	10	x		x or x	
	33	x		x or x	
	200		x	x or x	
1/f	10	x		x or x	
	33	x or x		x or x	
	200		x	x or x	

CONCLUSIONS

The computer program SAMSON is a useful tool for many purposes. The property of SAMSON to generate stationary noise with zero mean on the basis of an input (or standard) autocovariance function allows significant use of prior knowledge about analytical problems. The results presented above indicate that the agreement of the input ACVF with the resulting ACVF of the generated noise is good. This implies that it is possible to simulate noise with the same statistical properties as the noise of some source measured in practice. Consequently, software packages for many purposes can be tested exhaustively in circumstances giving a reliable picture of analytical practice. The example of testing a peak-finding procedure under different noise conditions indicates that the simulation of different kinds of noise to test the procedure is worth the extra effort: significant differences in performance with respect to different types of noise are observed. It would be useful to develop a more sophisticated test procedure for peak-finding procedures, because such a test procedure allows careful examination of the existing procedures and permits an optimal choice by the user for particular laboratory purposes.

Applications of SAMSON are possible in several fields. In information extraction methods, it can be used for simulation of detector signals or spectra interpretation. In data processing, it serves in filter procedures (e.g., Kalman filtering, forecasting). In systems theory, it may be useful in studies of the influence of different noise types on the output of a system and on the possibility of controlling a process, if the noise is applied at the input. In quality control, it may help in the development of test procedures to measure objectively the performance of software packages for information extraction. The steadily increasing computing and memory capacity of microcomputers allows development of more sophisticated methods for extraction of analytical information at low cost. The development of many of these methods may be aided by SAMSON.

REFERENCES

1 H. C. Smit and H. L. Walg, Chromatographia, 8 (1975) 311.
2 R. Deutsch, System Analysis Techniques, Prentice Hall, Englewood Cliffs, NJ, 1969, Ch. 5.
3 H. C. Smit and H. L. Walg, Chromatographia, 9 (1976) 483.
4 M. Abramowitz, Handbook of Mathematical Functions, Dover, New York, 1970, pp. 228, 231.
5 M. S. Bartlett, J. R. Stat. Soc., 138 (1946) 27.
6 G. M. Bobba and L. F. Donaghey, J. Chromatogr. Sci., 15 (1977) 47.
7 R. D. B. Fraser and E. Suzuki, Anal. Chem., 41 (1979) 37.
8 J. S. Milton and J. J. Corbet, Applied Statistics with Probability, Van Nostrand, London, 1979, p. 308.