J. Pharm. Pharmac., 1974, **26**, 629–636 *Received January* 29, 1974

Computer analysis of multicomponent ultraviolet spectra

B. W. MADSEN*, D. HERBISON-EVANS† AND J. S. ROBERTSON*

* *Pharmacy Department, University of Sydney, Sydney, Australia.*
† *Basser Computing Department, University of Sydney, Australia.*

In the assay of multicomponent systems by ultraviolet spectrophotometry, it is possible to improve accuracy and precision and also to obtain various statistical criteria by considering data over the range of wavelengths concerned. A linear least squares approach has been programmed to analyse mixtures of drugs in the presence or absence of known background. When the background was unknown but could be assumed of simple form, a non-linear alternative was successful.

Assay of pharmaceutical dosage forms may be performed by measurement of a multicomponent ultraviolet spectrum after dissolution and filtration and also by physical separation followed by individual component assays.

An obvious advantage of the multicomponent spectrum approach is simplicity. The process is easily learnt and there is less chance of something going wrong. Generality is another feature, meaning that new techniques with certain specific conditions need not be catered for with different preparations. Since there are less preparative steps involved than with many other methods, inaccuracies and assumptions have less chance of being propagated into serious errors. Associated with this is the favourable number of samples that can be processed in a given time. The pure spectra need only be determined once and placed in a library, reducing experimental effort still further. The apparatus is simple, with most quality control laboratories possessing an ultraviolet spectrophotometer, but not necessarily a gas chromatograph, spectrophotofluorimeter or polarograph. An important advantage is the ease of automation. There are various data logging systems now available (Larsen, 1973) with paper tape output which can be fed into a central computer error free. Even better, with the advent of the mini-computer, an on-line system can be contemplated.

Consequently there has been a good deal of interest in this type of assay. By using data over the whole wavelength range, Lübbers & Wodick (1969) were able to determine to approximately $\pm 1\%$ the individual concentrations of mixtures of four nucleotides in biological material. Glenn (1963) described the use of orthogonal polynomials to correct for irrelevant absorption in two component analysis, and since then there have been many reports of this technique in the pharmaceutical literature (Wahbi & Farghaly, 1970; Abdine, Wahbi & Korany, 1971, 1972; Wahbi & Abdine, 1973). With such an approach there is the advantage that the parameters enter linearly into the mathematics and so can be included in the linear regression.

However, with the non-orthogonal least squares approach adopted in this paper, it was found that polynomial representation of known background was not good, and a Gaussian curve was finally selected for this purpose.

Theory

In an ideal case, the background is assumed or found to be negligible, so that the predicted absorbance at any wavelength can be estimated by equation 1:

$$Y_{1\,pred} = a_0 + a_1X_{1l} + a_2X_{2l} + \ldots a_j\,X_{jl} \quad .. \quad .. \quad (1)$$

where a_j are the unknown concentrations of each of the j active ingredients and X_{jl} are the known pure spectra for these same components. The sum of squares of the deviations between the given ($Y_{1\,obs}$) and predicted absorbance values ($Y_{1\,pred}$) over all the wavelengths used can be minimized analytically from a developed set of linear simultaneous equations, giving the required concentration terms a_j. The a_0 term should be zero if the experimental technique is perfect, but its inclusion in the regression was always found in practice to improve the accuracy and precision of the concentration estimates (as will be shown later).

It should be realized that equation 1 can still be used when the background (diluent, lubricant, binder, preservative or whatever) spectrum is known. This means there will be one more term in equation 1 which allows for such background, but for which there will not be the same interest in the concentration estimate.

For dosage forms where the background is a mixture of excipients, there may be variation in the relative proportion of each of these additives from batch to batch, in addition to variation in the total amount. Where this applies, use of a fixed spectrum (X_{jl}) could lead to errors. With a variety of practical cases over a small wavelength range, it was found that the background was unimodal and could be adequately characterized by part of a Gaussian curve. The parameters of this curve produce non-linear equations which must be solved by a slower iterative process but then one may bypass ever having to determine experimentally the background spectrum.

Computation details

The linear equations were solved by matrix inversion using the double precision form of a standard IBM subroutine (MINV).* With approximately fourteen significant decimal digits, there was no evidence for concern over rounding errors with even the most poorly resolved systems and up to fifth order matrices. For the non-linear function minimization, a subroutine FUNMIN was written based on the Adaptive Simplex approach of Nelder & Mead (1965). Concentration standard deviations were estimated by the method of Kendall & Stuart (1961). The computed vector of parameter uncertainties was a sensitive indicator of approaching matrix singularity, so that long before significant errors occurred in the concentration estimates, they would be rejected because of high predicted relative standard deviations.

The Durbin-Watson statistic for serial correlation of residuals was estimated as described in Christ (1966) using equation 2:

$$d = \frac{\sum_{i=2}^{n} (\epsilon_1 - \epsilon_{1-1})^2}{\sum_{i=1}^{n} \epsilon_1^2} \quad .. \quad .. \quad .. \quad .. \quad (2)$$

* IBM System/360 Scientific Subroutine Package, 1967.

where ϵ_i is the ith predicted residual and n is the sample size. The distribution of d is symmetrical with a mean of 2 and rejection levels depend on the number of residuals and the number of parameters in the system. Inspection of equation 2 shows that if many neighbouring residuals are of the same sign (defined as positive serial correlation) then the sum of these terms will be small, and d will be small. In this work a one-tailed test for positive serial correlation only was used.

An IBM 7040 computer was used for the calculations and the program was written in FORTRAN IV.†

MATERIALS AND METHOD

Sulphacetamide (I), sulphadimidine (II) and sulphathiazole (III) were recrystallized from ethanol—chloroform to give melting points of 177–9°, 197–9° and 170–2° respectively. Colour tests and infrared spectra agreed with those listed in Clarke (1969), confirming identity. Acacia and lactose were of B.P. standard, while all other chemicals were of reagent grade. Spectrophotometric measurements were made with a Perkin-Elmer Model 124 double beam instrument and recorder using 1 cm quartz cells. Absorbance readings of the pure solutions and mixtures were taken at 2 nm intervals in the range 314–232 nm, unless otherwise indicated. These readings were automatically corrected for zero error between cells by the program.

For the first set of experiments, three one litre solutions in 0·05N sodium hydroxide were prepared, containing 13·35, 12·46, and 13·2 mg of I, II and III respectively. These solutions were then accurately mixed in all meaningful permutations of the following proportions:— 3,3,4; 2,4,4; 1,4½,4½; 2½,2½,5; 1½,1½,7 and 1,3,6 giving 21 mixtures for subsequent analysis, for which the individual component concentrations were known accurately. To rapidly test the performance of the program in resolving active ingredients when in the presence of unknown background, a wide range of possible excipient backgrounds was drawn by hand on graph paper (Fig 2). For each case, the absorbances were then read off and numerically added into the 21 mixture spectra.

For the second set of experiments, 3g of lactose and 750 mg of acacia were included in 3 litres of 0·05N sodium hydroxide. From this, three one litre solutions containing respectively 12·1, 11·6 and 11·7 mg of I, II and II were made. Ratios for these three solutions, in all meaningful permutations were 3,3,4 and 2,2,6 giving 6 synthetic mixtures.

In order to condense the numerous assay results and enable objective comparisons of different approaches to be made, the following procedure was adopted. The fractional error was given by equation 3:

$$\text{Fractional error} = \frac{\text{predicted concn} - \text{true concn}}{\text{true concn}} \times 100 \qquad (3)$$

For a given component in a series of assays, the variation of the fractional error about a mean was expressed as a standard deviation. The mean fractional error and its standard deviation were then combined to give the total error below.

$$\text{Total error} = \frac{\text{absolute value of mean}}{\text{fractional error}} + 2 \times \frac{\text{standard}}{\text{deviation}} \qquad (4)$$

† A listing is obtainable on request.

McFarren, Lishka & Parker (1970) have previously recommended a similar criterion for single valued concentration assays.

RESULTS AND DISCUSSION

The spectra of the three sulphonamides are shown in Fig. 1. They were chosen for the initial program development because they presented a poorly resolved situation; the assumption of negligible chemical interaction necessary for absorbances

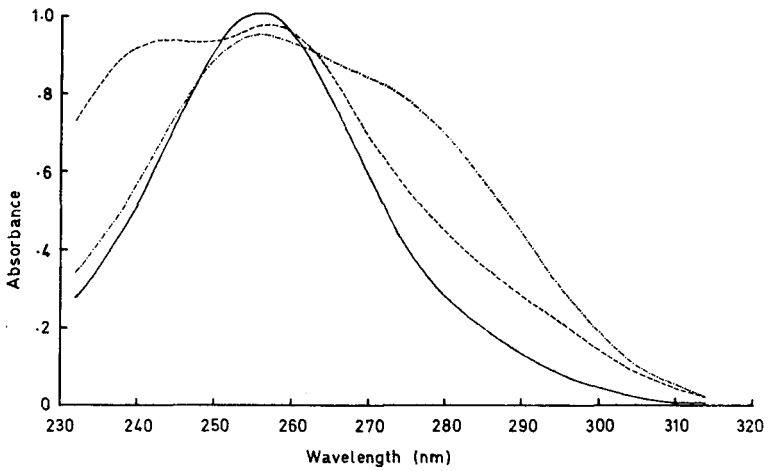

FIG. 1. Absorption spectra of the three sulphonamides in 0·05N sodium hydroxide. ——— sulphacetamide, 13·35 mg litre^{-1}: – – – – sulphadimidine, 12·46 mg litre^{-1} and —·— sulphathiazole, 13·2 mg litre^{-1}.

Table 1. *Assay errors using different approaches over 21 solutions.* Rows 1 and 2 show that generally the least squares procedure produces better estimates than the single point method. Rows 3 and 4 compare the two methods for robustness in the presence of spurious background. Rows 5 and 6 compare estimates when a blank term is included and absent respectively, over a wider wavelength range. When a_0 was included in the range 240–272 nm, the high total errors shown in Row 7 were obtained. The last row shows results for a linear solution when the added background was known (curve B in Fig. 2).

	Method	Constant term in regression	Total errors in percentages: Sulphacetamide	Sulphadimidine	Sulphathiazole	Wavelength range (nm)
1	Least SQ	No	7·3	14·5	3·8	240–272
2	Unique	—	13·2	18·9	5·1	230, 256, 280
3	Least SQ	No	15·4	34·4	10·3	240–272
4	Unique	—	23·0	51·0	8·1	230, 256, 280
5	Least SQ	Yes	7·1	9·9	2·5	232–314
6	Least SQ	No	16·0	17·4	4·6	232–314
7	Least SQ	Yes	39·3	63·4	46·8	240–272
8	Least SQ	Yes	4·4	3·7	4·0	232–314

of pure spectra to be additive was more likely to be valid for these similar species and because the problem had some pharmaceutical relevance.

Rows 1 and 2 of Table 1 compare the total error over the 21 solutions for the least squares method using equation 1 without the a_0 term in the range 240–272 nm with the three point approach using absorbances at 230, 256 and 280 nm. The relation of the total errors in Rows 1 and 2 to the more familiar mean fractional error and standard deviation is shown in Table 2. Notice that for two of the components, the total error has virtually been halved, and in the other it is significantly reduced. On this basis alone, the least squares approach considering 17 pts instead of 3 could probably be justified for routine analysis in quality control.

Table 2. *The contribution of accuracy and precision components to the total errors listed in Rows 1 and 2 of Table 1.* All figures are percentages.

Method	Sulphacetamide	Sulphadimidine	Sulphathiazole
Least squares			
Total error	7·3	14·5	3·8
Mean fractional error	—2·7	4·1	1·0
Standard deviation	±2·3	±5·2	±1·4
Unique			
Total error	13·2	18·9	5·1
Mean fractional error	— 3·2	2·3	1·9
Standard deviation	± 5·0	± 8·3	±1·6

However, there are other features which also support the current approach. Concentration uncertainties are calculated so that the analyst can quote results with 95% confidence intervals. With a unique solution, the analyst has to accept the results without ever knowing whether he has gone too far, or not far enough. Also with a unique solution the total error is dependent on the analyst successfully choosing the best wavelengths. While this may be easy for two well resolved spectra, it becomes increasingly difficult as resolution deteriorates and the number of components increases. With the least squares approach, no such decision is necessary as data are taken uniformly over the whole range. Another feature in favour of the present technique is that recording or transcription errors can be detected as commonly they will be outside 2–3 error standard deviations. With the simpler treatment, no such check is possible. Finally the least squares regression is also a semi-qualitative test. If there is additional background absorbance somewhere in the range, then the fit is not good, concentration uncertainties increase, and the probability that the regression residuals will be random decreases. The Durbin-Watson statistic is a quantitative measure of this randomness, and allows acceptance or rejection in the usual probability statements.

With more information on the spectra, one would expect that a least squares regression would produce concentration estimates more stable to non-ideal situations. Rows 3 and 4 in Table 1 show the effect of not correcting for zero error in the mixture spectra (less than ± ·005 absorbance units), bearing out this contention. Nevertheless, the large increase in total error does stress how important it is that the background be accurately known.

The best conditions found for assay of these data were with a wavelength range of 232–314 nm and the inclusion of the a_0 term in equation 1. Row 5 shows the sub-

stantial reduction in total error achieved for the component hardest to estimate, sulphadimidine. For comparison, the data in Rows 6 and 7 were included. Together with Rows 1 and 5, they show that inclusion of a blank term (a_0) was only beneficial when the whole spectrum was covered, and that without such a term, it was better to use only data with high signal to noise ratio (240–272 nm). With these findings, subsequent work was done from 232–314 nm with a_0 included in the regressions.

Fig. 2 shows the various backgrounds added into the mixture spectra for the computer simulations. Curve B was selected to illustrate the case of assaying for these sulphonamides in the presence of a known background. There were five unknowns in this linear regression; the three sulphonamides, the a_0 term and the background concentration. The last two were of no real interest, but their inclusion improved the sulphonamide estimates considerably (last row, Table 1).

Iterative estimates found in the presence of the backgrounds from Fig. 2 are summarized as total errors in Table 3. Rows 1, 2 and 4, 5 show that, as expected, the smaller the background to signal ratio, the better the Gaussian approximation;

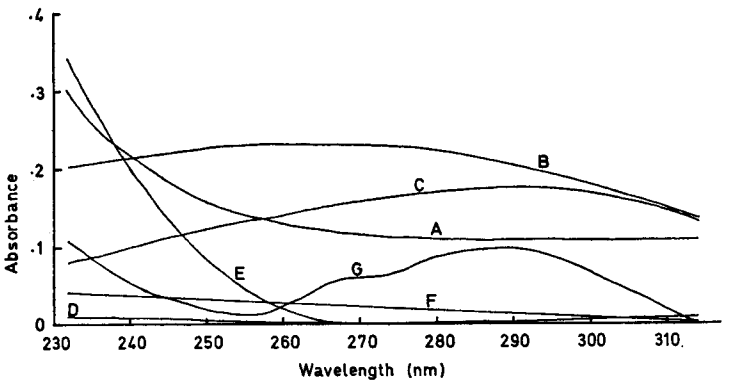

FIG. 2. Hand drawn backgrounds which were included in the mixture spectra for the computer simulations testing program performance. Curve F was an exact straight line. Curve G was included to show the consequences if the background cannot be adequately represented.

Table 3. *Total error in the determination of each of the sulphonamides in mixtures with varying backgrounds.*

	Background curve	Total error as percentage		
		Sulphacetamide	Sulphadimidine	Sulphathiazole
1	A	12·6	18·9	8·0
2	$\frac{A}{10}$	7·1	13·6	5·9
3	B	11·1	7·2	9·9
4	C	25·2	47·3	46·5
5	$\frac{C}{10}$	5·0	11·4	6·5
6	D	18·9	34·3	14·2
7	E	5·5	39·1	21·8
8	F	5·9	12·6	8·0
9	G	57	200	152

and therefore the better the active ingredient estimates. However, it is not necessary to fix arbitrary levels of signal to background ratio for good estimates, as these levels depend on many factors, and a final decision on assay feasibility is probably best made by the computer. For the systems studied so far, and within the error levels presented, there are adequate checks to prevent acceptance of invalid estimates.

Curve E illustrates the case with a large background absorbance, but since it is very dissimilar to the pure species, it can still be accurately allowed for. The results for Curve F show that under favourable circumstances, the program can even allow for an inclined perfect straight line. The last curve is atypical, but was included to show that the assumption of a unimodal background for one that is in fact more complex leads to intolerable errors. Needless to say, the investigator is warned not to accept these estimates.

The number of estimates accepted to those rejected can be made a compromise, based on the maximum assay error tolerated. For example, Table 4 presents the improved total errors attained with the data of Table 3, by discounting the whole regression when the probability of the residuals being random was less than 5%, and when individual concentration estimates were rejected if their relative standard deviations were greater than 5%. These results imply that even when an individual component in a mixture cannot be estimated with this approach, it does not necessarily jeopardize accurate assay of the other components.

Table 4. *The total errors can be improved by adjusting acceptance levels and rejecting those regressions and concentration estimates predicted outside these limits.* In the first row, 16 of the 21 regressions were considered acceptable on residual analysis, and of the remaining 48 concentration estimates, 36 had predicted relative standard deviations of less than 5%.

Background curve	Number of regressions accepted	Number of estimates accepted	Total error as percentages: Sulphacetamide	Sulphadimidine	Sulphathiazole
A	16	36	12·9	11·1	2·9
$\frac{A}{10}$	21	50	4·9	8·2	2·6
B	17	31	6·2	5·3	6·3
C	1	3	—	—	—
$\frac{C}{10}$	14	33	3·7	7·6	3·5
D	13	25	8·7	8·0	8·8
E	10	17	5·1	15·9	17·3
F	19	49	4·7	7·8	6·9
G	0	0	—	—	—

Total errors for the lactose-acacia background, together with the average execution time per mixture for the IBM 7040 are shown in Table 5. Notice that when the non-linear approach is used, the background need never be determined and that concentration estimates are more reliable, but that these two advantages are at the cost of increasing processing time.

Table 5. *Comparison of total errors and execution times with the lactose-acacia background for the non-linear and linear treatments.*

	Total errors as percentages			
	Sulpha-cetamide	Sulpha-dimidine	Sulpha-thiazole	Execution time per mixture
Linear (background known)	7·3	22·6	16·3	18 s
Non-linear (background unknown Gaussian approximation)	6·5	5·4	5·8	56 s

Additional points

The program was arranged to routinely print out the sum of the squared residuals at each wavelength for all the solutions treated in the batch, so that an error in one of the pure spectra is magnified over all the solutions, and very easiliy traced. For a valid series of regressions, this column should be approximately uniform over the whole range. If this is the case, it would imply that for the concentrations assayed, chemical interaction does not affect the ultraviolet spectra. The lower wavelength limit for absorbance reproducibility or instrument response linearity is also readily found from this column. A contour diagram showing the sum of squares versus the two non-linear variables is also printed out for non-linear regressions, enabling a decision on whether the minimum is unique or false.

REFERENCES

ABDINE, H., WAHBI, A. M. & KORANY, M. A. (1971). *J. Pharm. Pharmac.*, **23**, 444–447.

ABDINE, H., WAHBI, A. M. & KORANY, M. A. (1972). *Ibid.*, **24**, 518–521.

CHRIST, C. F. (1966). *Econometric Models and Methods*, p. 525. New York: Wiley.

CLARKE, E. G. C. (1969). *Isolation and Identification of Drugs*, London: The Pharmaceutical Press.

GLENN, A. L. (1963). *J. Pharm. Pharmac.*, **15**, *Suppl.*, 123T–130T.

KENDALL, M. G. & STUART, A. (1961). *The Advanced Theory of Statistics*, Vol. 2, p. 75. London: Griffin.

LARSEN, D. G. (1973). *Analyt. Chem.*, **45**, 217–220.

LÜBBERS, D. W. & WODICK, R. (1969). *Appl. Optics*, **8**, 1055–1062.

McFARREN, E. F., LISHKA, R. J. & PARKER, J. H. (1970). *Analyt. Chem.*, **42**, 358–365.

NELDER, J. A. & MEAD, R. (1965). *Computer J.*, **7**, 308–313.

WAHBI, A. M. & ABDINE, A. M. (1973). *J. Pharm. Pharmac.*, **25**, 69–72.

WAHBI, A. M. & FARGHALY, A. M. (1970). *Ibid.*, **22**, 848–850.