Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

ON THE CORRESPONDENCE BETWEEN WALSH SPECTRAL ANALYSIS AND 2^k FACTORIAL EXPERIMENTAL DESIGNS

Paul J. Sanchez
Systems and Industrial Engineering
University of Arizona
Tucson, AZ 85721, U.S.A.

ABSTRACT

This paper discusses the basic concepts of using a spectrally based approach to identify a polynomial response surface model. We show that this approach is closely related to classical experimental design methods. We concentrate on the use of Walsh spectra, and show that an experiment which is constructed in this fashion corresponds to a traditional 2^k factorial design.

1. BACKGROUND

Spectral analysis has its roots in Fourier analysis. It is essentially a method for examining periodicities in a set of data.

Consider a deterministic process y(t). Regardless of whether y(t) is a deterministic or a random process, the fundamental theorems of Fourier analysis state that as long as y(t) satisfies a set of requirements known as the Dirichlet conditions, it can be approximated arbitrarily closely on a finite interval by a linear combination of sine and cosine terms in the following fashion:

$$y(t) \approx \mu + \sum_{j=1}^{\infty} [A_{1,j}\sin(\omega_j t) + A_{2,j}\cos(\omega_j t)]. \quad (1)$$

This representation is called the Fourier transform of y(t), and the A's on the right hand side are called Fourier coefficients.

Suppose we take a finite sample of N observations of y(t) at evenly spaced discrete points which we will designate t = 1, 2,..., N. We denote this discrete sample as the set of observations $\{y_1, y_2, \ldots, y_N\}$. Such a sample can be modelled exactly as

$$y_t = \mu + \sum_{j=1}^{N/2} [A_{1,j}\sin(\omega_j t) + A_{2,j}\cos(\omega_j t)]. \quad (2)$$

where $\omega_j = 2\pi j/N$ for j = 1,...,N/2. Equation (2) is the finite Fourier transform representation of $\{y_t\}$. The process of estimating the coefficients for equation (2) is called Fourier analysis.

Walsh functions are a set of discrete-valued functions which assume only the values {-1,+1}. They are orthogonal and complete, having both even and odd symmetry. Each function is defined by convention on a fixed interval [0,T], and is written WAL(n,t) with $t \in [0,T]$. T is usually assumed to be 1, but can be any value if the function is scaled appropriately. The value n is an index which corresponds to the number of zero crossings on the interval [0,T], which is called the sequency of the Walsh function in an analogy to frequency in trigonometric functions. Walsh functions are paired by even and odd symmetry and referred to as CAL and SAL functions, respectively. These are defined as:

$$\left.\begin{aligned} CAL(k,t) &= WAL(2k,t) \\ SAL(k,t) &= WAL(2k-1,t) \end{aligned}\right\} k = 1,\infty \quad (3)$$

where k is the sequency.

The Walsh representation for a process {y(t)} looks much the same as the Fourier transform presented in equation (1). It is given by:

$$y(t) = \mu + \sum_{n=1}^{\infty} [B_{1,n} SAL(n,t) + B_{2,n} CAL(n,t)] \qquad (4)$$

where $B_{1,n}$, and $B_{2,n}$ are the Walsh coefficients for the odd and even functions, respectively.

Our main interest is to study discretely sampled data, so at this point we must make the transition from continuous Walsh functions into the discrete time domain. This is done by scaling the time axis relative to the highest order Walsh function of interest, and sampling the value of the continuous function over unit intervals. The result is a vector of N numbers which correspond to sampling WAL(k,t) at N equal intervals, i.e., at spacings of T/N. We will specify the discrete version as WAL(k,i), where i is the integer portion of [(Nt/T)+1], and define the value to be $WAL(k,i) \equiv WAL(k,t)$ for i = 1,...,N. WAL(k,·) will be used to indicate the vector consisting of WAL(k,i) for i = 1,...,N.

The value of N should be chosen so that each Walsh function has a unique vector associated with it. If k is the largest sequence number we wish to observe then we set $N = 2^{\lceil \log k \rceil}$, where the log is base 2 and $\lceil x \rceil$ is the smallest integer greater than x. As an example, if we were interested in Walsh functions up to order 5, we could represent them as vectors of length 8 whose elements assumed the values ±1. Two examples would be

WAL(0,·) = (1, 1, 1, 1, 1, 1, 1, 1),
WAL(5,·) = (1,-1,-1, 1,-1, 1, 1,-1).

Walsh functions are extremely easy to generate. The reader who wishes to see details of their construction is referred to Beauchamp (1975).

Arithmetic operations on Walsh functions make extensive use of the dyadic sum operator, which we will represent by "^". The dyadic sum is a bitwise XOR operation, where (p XOR q) is defined by the following table.

q \ p	0	1
0	0	1
1	1	0

Examples of the "^" operator are

$$\begin{array}{r} 7 \\ \hat{}\ \ 5 \\ \hline 2 \end{array} \iff \begin{array}{r} 111 \\ \hat{}\ \ 101 \\ \hline 010 \end{array} \quad \text{and} \quad \begin{array}{r} 9 \\ \hat{}\ \ 3 \\ \hline 10 \end{array} \iff \begin{array}{r} 1001 \\ \hat{}\ \ 0011. \\ \hline 1010 \end{array}$$

The "^" operator is used to define the Walsh multiplication property:

$$WAL(i,t)WAL(j,t) = WAL(i\hat{}j,t). \qquad (5)$$

Using the notation that N is the number of observations, y_j is the j^{th} term in the output time series, and Y_k is the k^{th} term in the transform series, the Walsh transformation is defined to be:

$$Y(k) = \frac{1}{N} \sum_{i=0}^{N-1} y_i WAL(k,i). \qquad (6)$$

In defining the Walsh transformation we are using the notation $y \mapsto Y$ to emphasize that the original observations and the estimated coefficients are a transform pair, either of which could be used to fully reconstruct the other. This notation is more concise than using $B_{1,j}$ and $B_{2,j}$, the Walsh coefficients which were defined in equation (4). We can change from one notation to the other using the relationships $Y_{2k} = B_{2,k}$ and $Y_{(2k)-1} = B_{1,k}$ for $k = 0,\ldots,\infty$.

The Walsh spectrum can be defined analogously to the Fourier spectrum. We will call the resulting estimator a Walsh periodogram, and denote it as P(s) where s is the sequency. P(s) can be estimated, in a manner analogous to the Fourier periodogram $I(\omega)$, by pairing the Walsh coefficients from equation (4) by sequency, and finding the sum of squares of each pair. This yields

$$P(0) = B_{2,0}^2$$

$$P(k) = B_{1,k}^2 + B_{2,k}^2$$

$$P(N/2) = B_{1,N/2}^2$$

for k = 1,2,...,(N/2)-1.

2. SPECTRAL EXPERIMENTAL DESIGN

We assume that the reader has some familiarity with the geometric viewpoint used by Scheffe (1959) to determine least squares estimators of linear models. The least squares estimators are found by orthogonal projection of the observation vector of "dependent" variables onto a sub-space defined by vectors of observations of "independent" variables. The independent variables are considered as potential explanatory or causal factors for the dependent variable. Using this viewpoint gives us a convenient perspective for discussing spectral estimators.

If we view the vector of dependent variables in a statistical linear model as a time series (in the sense that it is an indexed set), then the traditional Fourier spectrum represents the vector as a set of cyclic components which vary according to the time index. We know from the theory of Fourier analysis that any finite set of finite observations can be uniquely expressed as a set of Fourier coefficients, and the original observations can be reconstructed from those coefficients. The magnitude of the coefficient indicates the relative importance of a particular term in determining the outcome. In classical spectral analysis, terms with the same frequency are grouped together to give an estimator which is phase invariant. The periodogram is a plot of the resulting estimator versus frequency. It has the interpretation that a large value at a given frequency, often referred to as a "spike" in the spectrum, indicates that the factor which was varied at that frequency is an important component in the observed series.

Stated simply, we are finding a set of orthogonal functions, sines and cosines with integer multiples of frequencies. We then use it as a basis upon which we project the observed data, just as Scheffe does with linear models. This classical form of Fourier analysis is well understood, and is in common use by physical scientists because of the relatively intuitive interpretation of frequency when studying vibrations of matter, light, cycles in biological data, etc. However, sines and cosines are not the only function set which could be used as a basis. The procedure described above actually only used the completeness and orthogonality properties of sine/cosine functions. Recall that completeness means that there is no subspace which is orthogonal to the space spanned by our basis. The property of orthogonality gives us independence of the estimators and simplifies calculations. Any set of functions with these properties can be used to obtain a spectrum. The last fifty years has produced a large body of literature on alternate bases. There has been an upsurge of interest within the last twenty years in discrete function sets, such as Walsh and Haar functions, because of their applicability to digital signal processing.

Orthogonal decomposition has some nice properties for statistical analysis. Let us consider the form the analysis takes when Walsh functions are chosen. We assign a different Walsh function to each factor which is potentially in our model. The Walsh functions must be assigned so that if interactions exist, we have a uniquely identifiable Walsh function which corresponds to the interaction. This can be done (with some work) using the multiplication rule presented in the synopsis of Walsh properties. (An example is given later in the paper.) We then run our experiment by changing factor levels at time t according to the value of the assigned Walsh function at time t. The outcome of a Walsh function is either a +1 or a −1, so we should select an appropriate mapping for the factor being studied. For instance, if we are working with a continuous variable, we may choose to use the lowest factor setting of interest whenever the Walsh function is −1, and the highest value otherwise. Alternatively, we could do just the opposite. It is unimportant which mapping we choose as long as we are consistent.

We represent the observation series as a vector Y, and represent the basis set as a matrix W whose columns are Walsh functions.

Our goal is to find the vector of coefficients β such that $Y = W\beta$. If W is not a square matrix of full rank there is not a unique solution, but one solution is to use the least squares estimator for β, which is calculated using the formula $(W^tW)^{-1}W^tY$. With an orthogonal basis the computations are greatly simplified. Since the columns of W are orthogonal the W^tW term is a scalar times the identity matrix, and so is its inverse. In fact, it is easily verified that if W is an NxM matrix of Walsh functions the resulting least squares estimator for β is found by

$$\hat{\beta} = \frac{1}{N} W^tY.$$

Notice that this is the matrix form of the Walsh tranformation presented in equation (6). This means that the Walsh tranformation results in the least squares estimator for the vector β.

If we are willing to make the classical statistical assumption that the observations are independent normal random variables with common variance σ^2, i.e., the variance-covariance matrix is $\sigma^2 I$, then the variance-covariance matrix of $\hat{\beta}$ is calculated as follows.

$$\begin{aligned}
\mathrm{Cov}(\hat{\beta}) &= \mathrm{Cov}\left[\frac{1}{N} W^tY \right] \\
&= \frac{1}{N^2} W^t\mathrm{Cov}(Y)W \\
&= \frac{1}{N^2} W^t(\sigma^2 I)W \\
&= \frac{\sigma^2}{N^2} W^tW \\
&= \frac{\sigma^2}{N} I.
\end{aligned}$$

It follows immediately from the normality assumption that the estimators are independent.

Under the classical assumptions it can also be seen that the estimator consists of linear combinations of the observations and hence is itself normally distributed. Recall that our model is $Y = W\beta$. If we wish to test whether a given factor has a significant effect, we are testing whether the associated β is non-zero. For example, with Walsh vectors as a basis this is the same as stating a null hypothesis H_0 that there is no significant variation at sequency s if the corresponding $\beta_s = 0$. As shown by Sanchez (1983), squaring $\hat{\beta}_s$ and scaling by the variance yields a central χ^2 variable under the null hypothesis. If we wish to compare two different values, we can take the ratios of the squares of the β's. We saw above that the coefficients are independent, so the resulting χ^2 random variables will also be independent. Since the estimators all have a common variance the unknown variance terms cancel out when we take the ratio, and we obtain a valid F statistic.

Recall that the Walsh spectrum is estimated by summing the squares of pairs of Walsh coefficients. It follows immediately that this spectrum estimator consists of independent χ^2 random variables under the null hypothesis that there is no effect at a given sequency. This result can be used to do analysis of variance (ANOVA). If we could determine a priori that a particular sequency offers no contribution to the outcome, then the coefficient associated with that sequency could be attributed solely to the variance of the process. That coefficient could then be used as the denominator for our F tests in an ANOVA. Our problem is that we have an over-specified model. In Walsh analysis, as in Fourier analysis, all of the data points are fit exactly. There are no degrees of freedom left for estimating variance.

There is a solution to this problem. We can do as practitioners of ANOVA do, and assume based on prior knowledge or reasonable supposition that some of the terms do not actually belong in the model, and hence have zero coefficients. Using the reasoning given above, all such terms would constitute independent χ^2 variables, and because of their independence could be summed to increase the degrees of freedom for the denominator of the F test. Since we have chosen our function set for completeness, between them all of the estimated terms will account for all of the variability observed.

We partition the space into sub-spaces which we view as a factor space and an error space. In fact, this partitioning is a well known and important relationship called Parseval's theorem, and holds for Walsh analysis as well as Fourier analysis.

A second solution was proposed by Cogliano (1981). His recommendation was to make a "noise" run in which the factors are all held at nominal values, so that any variation observed in the output is attributable solely to the variance of the process being studied. Then make a second run in which factors are varied in a controlled fashion, and calculate the spectra for the two runs. Any large difference in the observed heights of spikes in the two spectra at the same sequency is due to factor effects.

3. EXAMPLE

We will use a 2^k factorial experiment as an illustration. In order to keep the example manageable, we will use k=3. Since we are only considering two factor levels for each of the 3 factors, Walsh functions are a natural choice. If we wish to have a full factorial design, we must choose the Walsh functions so that each factor and all possible interactions are identifiable. We will use the notation WAL(i,·) to denote a Walsh vector of appropriate length having sequence number i. Then if we assign WAL(i,·) to factor X_1 and WAL(j,·) to X_2, we would look for an X_1X_2 interaction at WAL(i,·)WAL(j,·). The product of two Walsh functions is another Walsh function, which can be identified using relationship (5).

The problem of selecting sequencies such that there will be no overlap of the identifying terms for interactions is discussed in detail in Sanchez (1985). For example, we would not in general wish to assign Walsh functions 1, 2, and 3 to factors 1, 2 and 3, respectively. If we did so, we would be unable to state whether a spike at WAL(3,·) was due to factor 3 or to an interaction between factors 1 and 2. However, for a full factorial design, one simple solution is to reserve one bit location for each of the main effect factors, i.e., each factor is assigned a Walsh function which is a power of 2. When there are no bit locations shared by the two numbers, a dyadic sum is the same as normal addition. Thus, if we have k factors we will use Walsh functions 2^0 through 2^{k-1}. For our example this means factor 1 is assigned WAL(1,·), factor 2 is assigned WAL(2,·), and factor 3 is assigned WAL(4,·). We get the following table for observing factors and effects.

Model Term	Observed At
X1	WAL(1,·)
X2	WAL(2,·)
X3	WAL(4,·)
X1X2	WAL(3,·)
X1X3	WAL(5,·)
X2X3	WAL(6,·)
X1X2X3	WAL(7,·)

This corresponds to the design matrix given in Table 1. The estimators would be obtained by taking the vector product of each column with the vector of observations, Y.

Actually, it is both equivalent and computationally easier to do a Fast Walsh Transform (FWT) to obtain these values. The statistical analysis would then consist of designating one or more of the terms as our error space and constructing F ratios. We would most likely use the three way interaction term, which would correspond to WAL(7,·) in this design tableau. We would construct the sample F ratios by squaring all of our Walsh estimators and taking the ratio of each one to the squared estimator for WAL(7,·). These ratios would be compared to the value of an $F_{1,1}$ with appropriate p-value in a table to see if we should accept or reject the hypothesis of no factor effect. If more than one term were designated as belonging in the error space, the denominator for our F statistic would be found by summing the squared terms, scaling by the number of terms being included, and adjusting the degrees of freedom appropriately in the F test. Overall degrees of freedom can be

Table 1: Design Matrix for a 2^3 Factorial Experiment

mean WAL(0,·)	X1 WAL(1,·)	X2 WAL(2,·)	X3 WAL(4,·)	X1X2 WAL(3,·)	X1X3 WAL(5,·)	X2X3 WAL(6,·)	X1X2X3 WAL(7,·)
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	-1	-1	-1	1	1
1	1	-1	1	-1	1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	1	1	-1	-1	1	-1

increased by replicating the experiment, i.e., making a longer experimental run so that each Walsh function is repeated several times.

4. CONCLUSIONS

The evaluation which results from using Walsh analysis is identical to a classical ANOVA 2^k factorial experimental design. We obtain the same design matrix which could be found in any classical textbook on experimental design, and end by constructing exactly the same F test. There are several reasons why this should interest us. First, it provides a new perspective on an old field, experimental design. I personally find it easier to design a 2^k factorial experiment from a sequency perspective than to resort to look-up tables for the design. (Walsh functions are extremely easy to generate, so a small program replaces a statistics text.) In addition, this perspective may lead to more insight in the area of experimental design. Secondly, the FWT provides a very efficient mechanism for doing the ANOVA calculations. It takes only $O(N \log_2 N)$ additions or subtractions to do a FWT on N observations. (Davies (1956) implicitly uses the FWT when describing how to evaluate a 2^k factorial experimental design by hand.) Finally, we can perhaps begin to see why the spectral approach to significant parameter identification used by Schruben, Cogliano, and Sanchez has succeeded empirically in producing results which are so consistent with run-oriented experiments.

ACKNOWLEDGEMENTS

The author wishes to thank Professor Lee W. Schruben of Cornell University, who originally conceived using spectral analysis for significant factor identification, for his continuing encouragement and support of this work. He would also like to thank his wife Susan for her assistance (and patience) in reading through numerous revisions.

REFERENCES

Beauchamp, K. G. (1975). Walsh Functions and their Applications. Academic Press, New York.

Cogliano, V. J. (1981). Sensitivity Analysis and Model Identification in Simulation Studies: A Frequency Domain Approach. Unpublished Ph.D. dissertation, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Davies, O. L. (1956). Design and Analysis of Industrial Experiments, Second Edition. Academic Press, New York.

Priestley, M. B. (1981). Spectral Analysis and Time Series. Academic Press, New York.

Sanchez, P. J. (1983). Walsh Functions for Spectral Analysis. In: Proceedings of the 1983 Winter Simulation Conference (S. Roberts, J. Banks, and B. Schmeiser, eds.). Institute of Electrical and Electronic Engineers, Washington, D.C., 405-406.

Sanchez, P. J. (1985). Significant Factor Identification Using Discrete Spectral Methods. Unpublished M.S. thesis, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Scheffe, H. (1959). The Analysis of Variance. Wiley, New York.

Schruben, L. W. and Cogliano, V. J. (1981). Sensitivity Analysis of Discrete Event Simulations: A Frequency Domain Approach. Technical Report 514, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

AUTHOR'S BIOGRAPHY

PAUL J. SANCHEZ is on the faculty of the Systems and Industrial Engineering department of the University of Arizona. He received an S.B. in economics from M.I.T. in 1977 while concurrently working as a statistical programmer at the Harvard School of Public Health. He then did econometric modelling and programming at the University of Rhode Island for a year. From 1978 to 1981 he worked for Mitrol, Inc., a company which specialized in manufacturing databasesand MRP. He did technical support, consulting, and R&D. He returned to academic studies in the Operations Research department at Cornell University in 1981, received an M.S. in 1985, and is expecting completion of his Ph.D. shortly. His current research interests include simulation output analysis, spectral analysis, and systems modelling.

Paul Sanchez
Systems & Industrial Engineering
University of Arizona
Tucson, AZ 85721, U.S.A.
(602) 621-6558