SOME TESTS OF THE VANÍČEK METHOD OF SPECTRAL ANALYSIS

J. TAYLOR and S. HAMILTON
Oceanographic Research Division, Marine Sciences Branch, Dept. of the Environment, Ottawa, Ont., Canada

(Received 20 March, 1972)

Abstract. One of the chief claims of the Vaníček Method of spectrum analysis is its ability to remove 'systematic noise' from a time series with minimal distortion of the spectrum of the remaining series. In this paper, comparisons are made between the Vaníček Method and the more usual method consisting of preliminary removal of the systematic noise followed by fourier analysis. Formulas relating the two methods are developed and a series of comparative plots of simple spectra are presented. Results indicate that the Vaníček Method does not displace a spectral peak falling close to the systematic noise, that it distorts the amplitude of the peak less than the conventional method of analysis, and that it tends to intensify the side lobes.

1. Introduction

In his paper Vaníček (1971) has described a new method of spectrum analysis. One of the chief claims of the method is its ability to remove 'systematic noise' from a time series with minimal distortion of the spectrum of the remaining series. By systematic noise, are meant such things as an unknown constant term (or datum), a linear trend of unknown size, or a trigonometric wave of known frequency, but unknown amplitude and phase. These are superimposed upon a series consisting of waves of various frequencies (and whose spectrum is required) plus perhaps random noise.

A commonly used method of dealing with a series containing a 'systematic noise function' consists in the preliminary removal of the 'noise function' by means of least squares fitting to determine its amplitude and then fourier analysis of the remaining series to determine its spectrum. This technique causes considerable degradation of the spectrum near the frequency of the noise function removed (datum and trend may be thought of as cosine and sine waves of zero frequency).

Vaníček's method combines the least squares fitting and fourier analysis in a novel way and has the property of preserving the position (though not the amplitude) of any single peak of the spectrum no matter how close in frequency to the interfering noise function. In this paper we will develop formulas for comparing Vaníček's spectral estimates with those of the more common method mentioned just previously, and we will present a series of comparative plots of simple spectra.

It should perhaps be mentioned here that the Vaníček method possesses other features, enabling it to deal with data with gaps and variable datum, unequally spaced data, etc. We will make no attempt to assess these features, but will deal only with uniformly spaced time series.

2. Mathematical Development

Suppose we have the time series:

$$F(t) = T(t) + K\phi(t), \tag{1}$$

where

$$T(t) = \sum_{j=1}^{m} a_j \cos f_j \pi t/T + b_j \sin f_j \pi t/T .$$

$F(t)$ is defined at n equally spaced values of time t in the range $t = -T$ to T. The frequencies $f_1 \ldots f_m$ are in cycles per $2T$ seconds. $\phi(t)$ is the systematic noise function. We assume $\phi(t)$ to be either an odd function of time or an even function, but not a combination. In particular we will deal here with four possible forms of $\phi(t)$ only, namely:

$$\begin{aligned} \phi(t) &= 1 \text{ (datum)} \\ &= t \text{ (linear trend)} \\ &= \cos f_\phi \pi t/T \\ &= \sin f_\phi \pi t/T \end{aligned} \quad \Big\} \; f_\phi \text{ is some known frequency.}$$

Given $F(t)$ as data, and the form of $\phi(t)$, but none of the frequencies f_j or the coefficients K, a_j, b_j, it is desired to remove the effect of $\phi(t)$ as well as possible from $F(t)$, and estimate the frequencies and amplitudes of the waves present in $T(t)$. (Actually we will deal with energy or variance rather than amplitude in order to be consistent with Vaníček's approach.)

We will use the following notation:

$(F, G) = \sum_t F(t) \cdot G(t) =$ Inner product of F and G,

$\|F\| = \sqrt{(F, F)} =$ Norm of F,

$R_{F,G} = (F, G)/\|F\| \cdot \|G\| =$ Correlation of F and G.

For simplicity we will denote $\sin f_j \pi t/T$ by $s_j(t)$ and $\cos f_j \pi t/T$ by $c_j(t)$. Thus the correlation between $\sin f_j \pi t/T$ and $\phi(t)$ for example is R_{ϕ, s_j}. If 'f' is some arbitrarily chosen frequency, we denote $\sin f\pi t/T$ by $s(t)$, and the correlation of this with $\phi(t)$ by $R_{\phi, s}$, etc.

We now proceed to work out and compare the variance estimates $\hat{V}(f)$ and $\hat{\hat{V}}(f)$ obtained using the 'least squares plus fourier analysis' technique, and the Vaníček technique respectively.

3. The Estimate $\hat{V}(f)$

We now calculate the estimate $\hat{V}(f)$ of the variance of $T(t)$ at an arbitrary frequency f, by first removing $K \cdot \phi(t)$ from $F(t)$ by least squares fitting, then applying fourier analysis to the remaining series.

First we calculate the least squares estimate of K:

$$\|\phi\|^2 \cdot \hat{K} = (\phi, F).$$

Next, we attempt to remove $\phi(t)$ from $F(t)$:

$$G(t) = F(t) - \hat{K}\cdot\phi(t).$$

Then we obtain $\hat{a}(f)$ and $\hat{b}(f)$, the fourier coefficients at any frequency f:

$$\|c\|^2\cdot\hat{a}(f) = (c, G),$$
$$\|s\|^2\cdot\hat{b}(f) = (s, G).$$

Finally the even and odd portions of the estimated variance at frequency f are obtained:

$$\hat{V}_E(f) = \hat{a}(f)^2\cdot\|c\|^2,$$
$$\hat{V}_O(f) = \hat{b}(f)^2\cdot\|s\|^2.$$

Working through the algebra we get:

$$\hat{V}_E(f) = \left\{\sum_{j=1}^{m} \|c_j\|\, a_j\,(R_{c,c_j} - R_{\phi,c}\cdot R_{\phi,c_j})\right\}^2,$$
$$\hat{V}_O(f) = \left\{\sum_{j=1}^{m} \|s_j\|\, b_j\,(R_{s,s_j} - R_{\phi,s}\cdot R_{\phi,s_j})\right\}^2. \qquad (2)$$

The total variance is the sum of the odd and even parts. (The $\hat{V}$'s defined here are actually 'n times' variances).

It may be noted that we allow f to be any arbitrary frequency. This is compatible with the Vaníček approach. However, the usual fourier transform computer routines give results only at whole multiples of the fundamental frequency (one cycle per $2T$). Actually a finer spacing of frequency may be obtained without too much effort by padding the data record out with zeros to several times its original length (see 3) and making simple corrections for the norms $\|c\|$ and $\|s\|$ at nonintegral frequencies. Of course, f should still be less than the cut-off frequency of half a cycle per time increment.

4. The Estimate $\hat{\hat{V}}(f)$ by Vaníček's Method

The Vaníček estimate of the variance of $T(t)$ at frequency f is defined by

$$\hat{\hat{V}}(f) = \|G\|^2 - \|H_f\|^2,$$

where, as before, $G(t)$ is the residual series left after $\phi(t)$ has been removed from $F(t)$ by least squares. $H_f(t)$ is the residual series after $\phi(t)$, $\cos f\pi t/T$ and $\sin f\pi t/T$ have been simultaneously removed from $G(t)$ by least squares.

Working through the algebra we get the following expressions for the even and odd parts of $\hat{\hat{V}}(f)$:

$$\hat{\hat{V}}_E(f) = \hat{V}_E(f)/(1 - R^2_{\phi,c}),$$
$$\hat{\hat{V}}_O(f) = \hat{V}_O(f)/(1 - R^2_{\phi,s}). \qquad (3)$$

5. Generalization of Results

Suppose now that there are several different noise functions in $F(t)$, that is

$$F(t) = T(t) + K_1\phi_1(t) + K_2\phi_2(t) + \cdots K_L\phi_L(t)$$

where, as before, each of the ϕ's is either an odd or an even function of time. It may be shown that Equations (3) still hold if $R_{\phi,c}$ and $R_{\phi,s}$ are now interpreted as 'multiple correlations' between $\cos f\pi t/T$ or $\sin f\pi t/T$ and the set $\phi_1(t)$, $\phi_2(t) \ldots \phi_L(t)$.

As our results depend upon the orthogonality of cosine and sine waves they may not be directly generalized to unequally spaced time series.

A. COMPARISON OF THE ESTIMATES $\hat{V}(f)$ AND $R(f)$

(i) It is interesting to note that $\hat{V}(f)$ and $\hat{\hat{V}}(f)$ are independent of K; that is, the amount of the function $\phi(t)$ present in $F(t)$ does not affect the results at all.

(ii) From the form of Equations (3) it is evident that the two methods will give almost identical results if $\phi(t)$ is only slightly correlated with $\cos f\pi t/T$ or $\sin f\pi t/T$. For example, if $\phi(t)=t$, then for all frequencies greater than the fourth harmonic, we find that $(1-R^2_{\phi,s})$ is greater than 96%.

Also, it is interesting to note from Equations (3) that the Vaníček estimate $\hat{\hat{V}}$ will always be larger than $\hat{V}$.

(iii) Suppose the series $T(t)$ contains just one frequency, namely f_1. From formulas (2) and (3) we find the estimates $\hat{V}$ and $\hat{\hat{V}}$ at frequency f_1 to be:

$$\hat{V}_E(f_1) = \|c_1\|^2\, a_1^2(1 - R^2_{\phi,c_1})^2, \qquad \hat{V}_O(f_1) = \|s_1\|^2\, b_1^2(1 - R^2_{\phi,s_1})^2,$$
$$\hat{\hat{V}}_E(f_1) = \|c_1\|^2\, a_1^2(1 - R^2_{\phi,c_1}), \qquad \hat{\hat{V}}_O(f_1) = \|s_1\|^2\, b_1^2(1 - R^2_{\phi,s_1}).$$

The true values at $f=f_1$ are:

$$V_E(f_1) = \|c_1\|^2\, a_1^2, \qquad V_O(f_1) = \|s_1\|^2\, b_1^2.$$

Thus, provided there is no interference from other frequencies, the Vaníček estimate of variance is the geometric mean of the true value and the estimate $\hat{V}$. If R_{ϕ,c_1} and R_{ϕ,s_1} are fairly small, this means that the Vaníček estimate is less biased than $\hat{V}$ by about 50%.

Also, if we evaluate the derivative $\mathrm{d}\hat{\hat{V}}/\mathrm{d}f$ at $f=f_1$ we find it to be zero, confirming that $\hat{\hat{V}}$ peaks at the correct frequency in the absence of interference. This is not generally true of $\hat{V}$.

(iv) If there is no noise function $\phi(t)$ present in $F(t)$ and no attempt is made to remove it, then the variance estimate, which we will call $\mathring{V}(f)$, may be obtained from Equations (2) as:

$$\mathring{V}_E(f) = \left\{\sum_{j=1}^{m} \|c_j\|\; a_j R_{c,c_j}\right\}^2,$$
$$\mathring{V}_O(f) = \left\{\sum_{j=1}^{m} \|s_j\|\; b_j R_{s,s_j}\right\}^2. \tag{4}$$

This is the form to which both $\hat{V}$ and $\hat{\hat{V}}$ reduce in the absence of any function $\phi(t)$, and we will use it as a kind of standard of perfection in our comparisons of $\hat{V}$ and $\hat{\hat{V}}$.

At this particular stage we wish to indicate how the Vaníček method affects sidelobes (in the region of $f=f_\phi$), from waves whose frequencies are some distance from f_ϕ. We cannot compare $\hat{\hat{V}}$ against $\hat{V}$, for this latter estimate goes to zero in the region of $f=f_\phi$, as may be seen from Equations (2). So we will compare against $\mathring{V}$. By going through a considerable amount of calculations it may be shown that the $\hat{\hat{V}}$ sidelobes are generally somewhat larger than the $\mathring{V}$ sidelobes in the region of f_ϕ. So we may say that the Vaníček Method somewhat magnifies interference from the various constituents of $T(t)$, especially in the region of f_ϕ. More specific information on this point will be given in the sections describing Figures 2, 3 and 4.

When $T(t)$ consists solely of the constituents of random noise, then Vaníček's method will produce (on the average) a spectrum which is neither reduced nor magnified in the region of f_ϕ. This surprising property may be proved using elementary probability. Evidently the degradation of peaks near f_ϕ is just compensated by the magnification of the side lobes from peaks at more distant frequencies.

(v) There is a third method of removing $\phi(t)$ from $F(t)$ which is of interest here for comparison purposes. Suppose we find the least squares fit to $F(t)$ using the three functions $\phi(t)$, $\cos f\pi t/T$ and $\sin f\pi t/T$ at each frequency f. From the coefficients of the latter two functions, the estimate $\hat{\hat{\hat{V}}}(f)$ may be obtained. The formulas for this estimate turn out to be:

$$\begin{aligned}\hat{\hat{\hat{V}}}_{\mathrm{E}}(f) &= \hat{\hat{V}}_{\mathrm{E}}(f)/(1-R^2_{\phi,c}),\\ \hat{\hat{\hat{V}}}_{\mathrm{O}}(f) &= \hat{\hat{V}}_{\mathrm{O}}(f)/(1-R^2_{\phi,s}).\end{aligned} \tag{5}$$

If $T(t)$ consists of one frequency f_1 only, then it is evident that $\hat{\hat{\hat{V}}}$ will give a perfect estimate of variance at f_1. However, it also turns out that this method severely magnifies the sidelobe interference in the region of f_ϕ.

So the Vaníček estimate $\hat{\hat{V}}$, which is evidently the geometric mean of $\hat{V}$ and $\hat{\hat{\hat{V}}}$, may be looked upon as a compromise mid-way between these two estimates.

We now proceed to illustrate the Vaníček method using a series of comparative plots of simple spectra.

B. FIGURE 1

Here we have plotted $\mathring{V}$, $\hat{V}$, $\hat{\hat{V}}$ and $\hat{\hat{\hat{V}}}$ against frequency f. $\phi(t)$ is a sine (or cosine) wave of fairly high frequency f_ϕ, which is indicated by the vertical dotted line. The series $T(t)$ consists of a single sine (or cosine) wave whose frequency f_1 is indicated by the position of the heavy vertical line. The variance of this wave is shown by the height of this heavy line. The spacing of the tick marks on the x, or frequency, axis is the elementary harmonic spacing; that is, $\Delta f=1$ cycle per $2T$ seconds. In Figure 1, f_1 is half a cycle from f_ϕ. As we are dealing with fairly high frequencies here, the left end of the frequency axis should not be thought of as $f=0$.

Before proceeding to compare the four plots, it might be worth while to summarize here the definitions of the four estimates of variance.

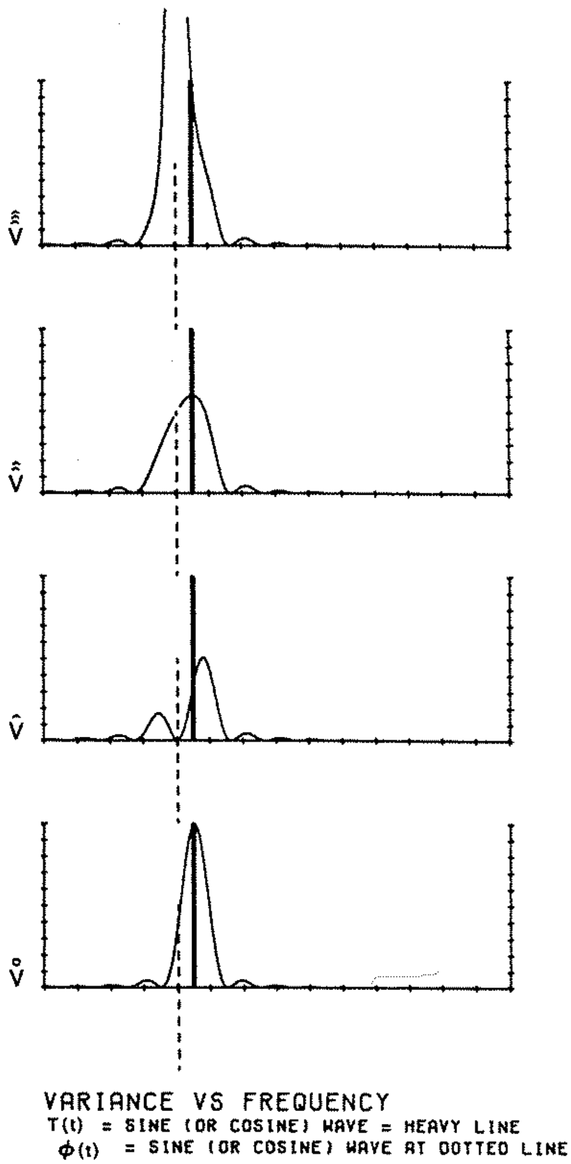


Fig. 1.

$\mathring{V}(f)$ – Straight fourier analysis of $T(t)$ is used, that is $\phi(t)$ is assumed to be absent from $F(t)$. This is the standard or ideal.

$\hat{V}(f)$ – $\phi(t)$ is removed by least squares, then fourier analysis is applied.

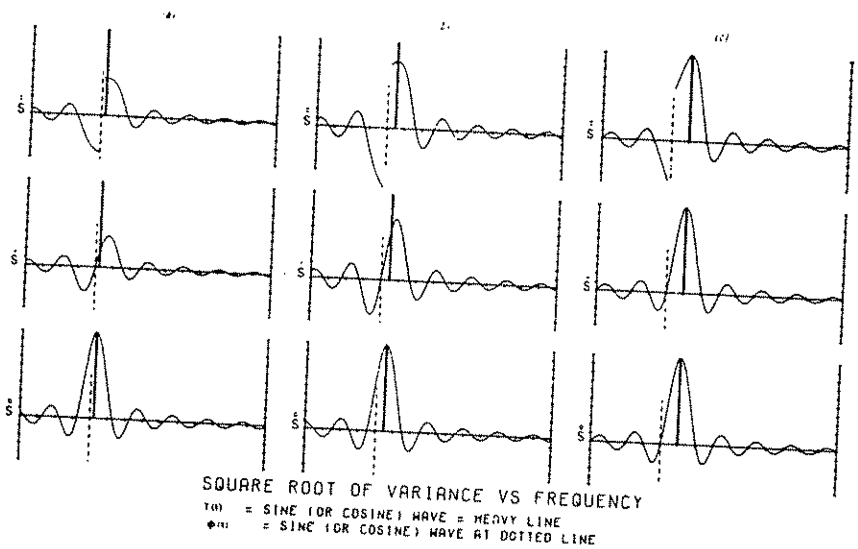
$\hat{\hat{V}}(f)$ – This is the Vaníček estimate.

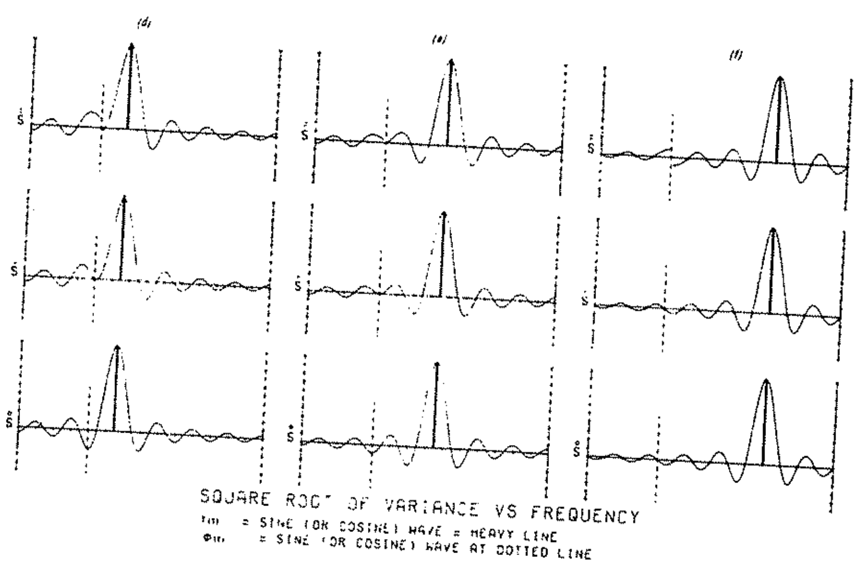
$\hat{\hat{\hat{V}}}(f)$ – At each frequency f, a least squares fit is made to $F(t)$ using $\phi(t)$, $\cos f\pi t/T$ and $\sin f\pi t/T$.

In Figure 1, we see that the Vaníček estimate remains finite at frequency f_ϕ, while $\hat{V}$ goes to zero and $\hat{\hat{\hat{V}}}$ goes to infinity. Also, we note that $\hat{\hat{V}}$ reaches its peak at frequency f_1 as expected. $\hat{\hat{V}}$ gives the correct value of variance at f_1, but this would be no help in the search for an unknown peak as may be plainly seen from the shape of the curve $\hat{\hat{V}}(f)$. The estimate $\hat{V}(f)$ shifts the peak a bit to the right of f_1 and introduces another smaller peak to the left of f_ϕ. $\mathring{V}(f)$ is of course the familiar $(\sin x/x)^2$ function. The Vaníček estimate $\hat{V}$ is evidently superior to both $\hat{V}$ and $\hat{\hat{V}}$.

C. FIGURES 2a–2f

These are similar to Figure 1 for a variety of values of f_1–f_ϕ. However, the estimate $\hat{\hat{V}}(f)$ has been dropped, and we have plotted the square root of variance against frequency, rather than the variance itself. We denote this new variable by $S(f)$ (for 'standard deviation'). Plots of $S(f)$ should be more useful than $V(f)$ in allowing one to make rough visual estimates of the interfering effect of sidelobes at frequencies near f_ϕ; for total variance may be obtained by superimposing various curves of form



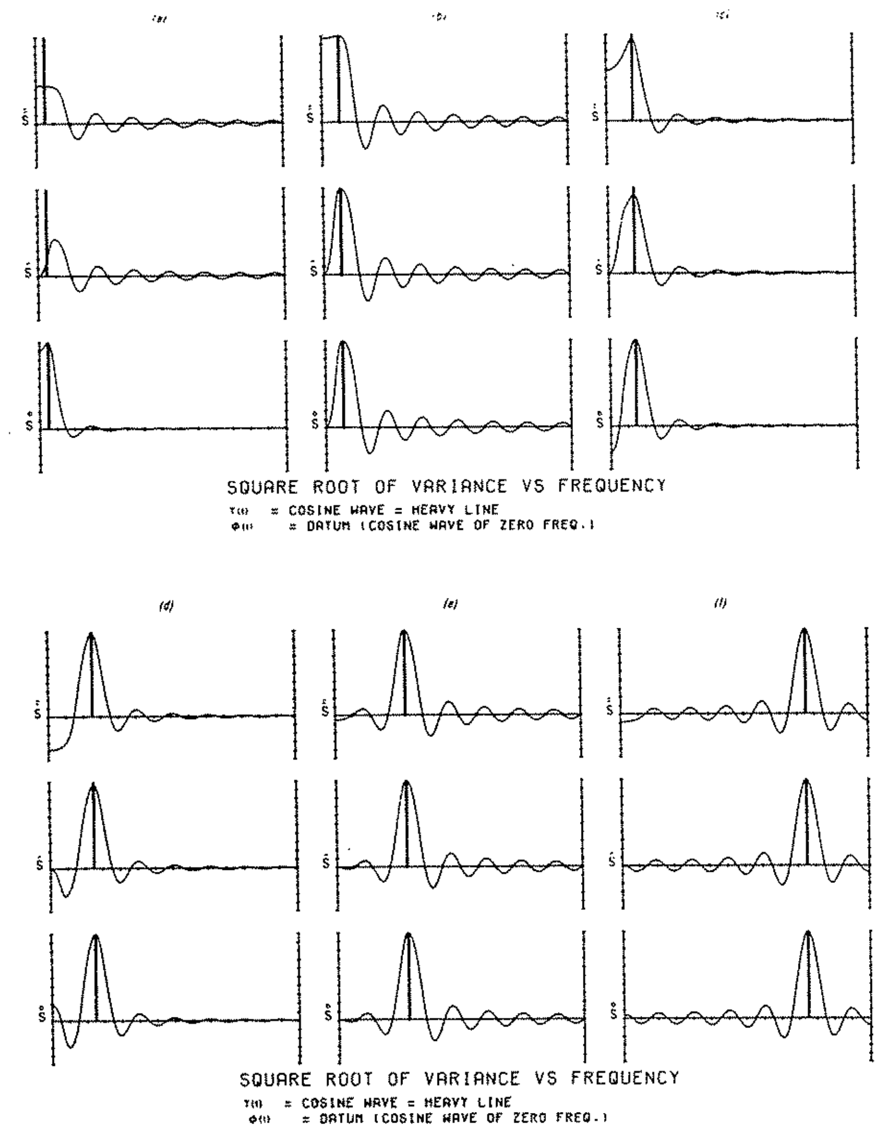


Figs. 2a–f.

$S(f)$ and squaring the results. For example, Figure 2f might be superimposed upon Figure 2b, in any desired proportion, to determine the effect of the sidelobes of a peak six cycles from f_ϕ, upon a peak half a cycle from f_ϕ.

It is interesting to note that the Vaníček estimate $\hat{\hat{S}}(f)$ changes sign abruptly at frequency f_ϕ. Yet if the portion of this curve to the left of f_ϕ were to be inverted, it would be found to flow smoothly into the right-hand part. The squaring to obtain variance will always eliminate this discontinuity.

From Figure 2f especially, it may be seen that the sidelobes of $\hat{\hat{S}}(f)$ near frequency f_ϕ are a bit bigger than those of $\mathring{S}(f)$. It may be shown that the Vaníček method magnifies sidelobes from distant peaks by a factor of $\sqrt{3}$, on the average, in the region of f_ϕ.



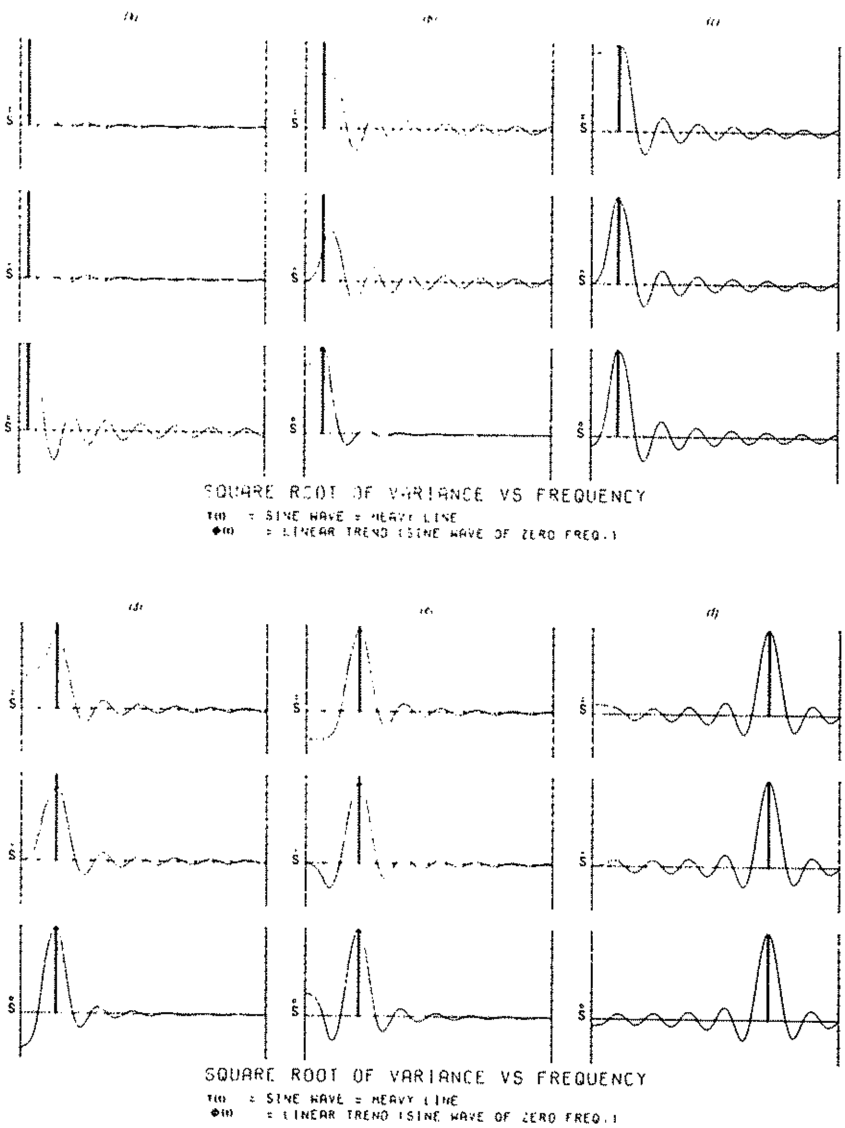
Figs. 3a–f.

D. FIGURES 3a–3f

Here $\phi(t) = 1$ = datum or constant term, and $f_\phi = 0$. The left hand end of these plots is at $f = f_\phi = 0$. We see that the Vaníček method still locates the peak variance at frequency f_1, as it should. However, at the lower frequencies this peak is quite flat, making its exact position somewhat hard to determine. Figure 3f indicates how the Vanicek method magnifies side lobes at low frequencies. Here the magnification factor is $\sqrt{5}$.

E. FIGURES 4a–4f

In this case $\phi(t) = t$ = trend, and $f_\phi = 0$. The results here are similar to those of Figure 3, and the flattening of low frequency peaks by the Vaníček method is especially notice-



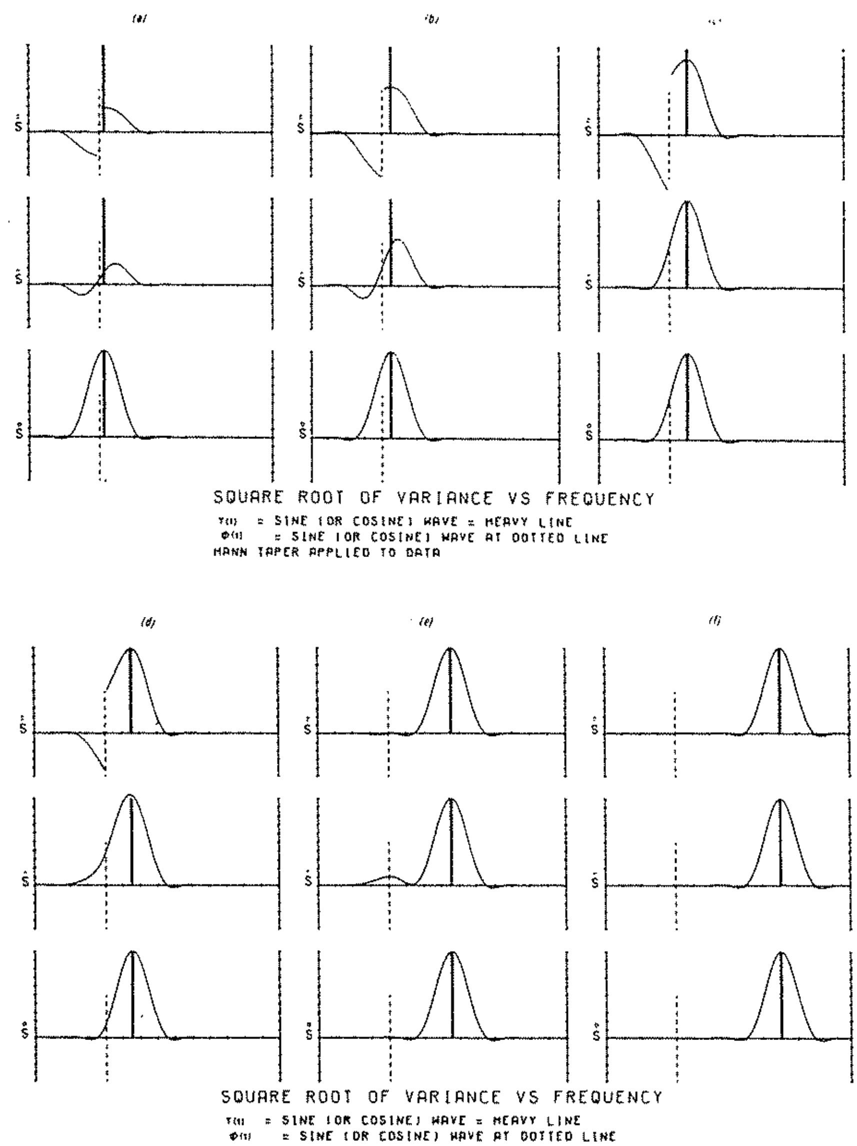
Figs. 4a–f.

able in Figures 4a, 4b, and 4c. However, these flat peaks would seem to be preferable to misleading estimates $\hat{S}$ in Figures 4a and 4b where the peaks maintain their sharpness, but are shifted to higher frequencies.

Figure 4f shows how the Vaníček method magnifies side lobes at low frequencies. In this case, the magnification factor is $\sqrt{7}/3$.

F. DATA TAPERING TO REDUCE SIDELOBES

As the Vanicek method does cause some magnification of sidelobes, it would seem to be especially worth while to find some way of incorporating data tapering into it. Vaníček (1971) discusses this in his paper but seems to come to no final conclusion.



Figs. 5a–f.

The procedure for calculating $\hat{V}(f)$ might be modified to include tapering as follows:

(i) Remove $\phi(t)$ from $F(t)$ by least squares.
(ii) Apply the taper (say $W(t)$) to the remaining series $G(t)$.
(iii) Apply fourier analysis to the resulting series $W(t)\cdot G(t)$.

We tried (among other things) adapting this technique to the Vaníček method, but got rather poor results. However, we did find one technique which worked pretty well. This consists in applying the square root of the taper to all functions entering into the analysis. That is, we apply the Vaníček technique exactly as before but employ the functions $\sqrt{W(t)}\cdot F(t)$, $\sqrt{W(t)}\cdot\phi(t)$, $\sqrt{W(t)}\cdot\cos f\pi t/T$ and $\sqrt{W(t)}\cdot\sin f\pi t/T$ instead of $F(t)$, $\phi(t)$, etc.

Figures 5a–5f are plots obtained using this technique for the Vaníček estimates $\hat{\hat{S}}(f)$, and the technique mentioned earlier for the estimates $\hat{S}(f)$. We used the Hann taper, $W(t)=\cos^2 t\pi/2T$. $T(t)$ and $\phi(t)$ are sine (or cosine) waves of fairly high frequency as in Figure 2. The Hann taper generally reduces all amplitudes to half their proper values, so we have doubled all results to simplify comparisons.

The plots show that the Vaníček method has maintained its fundamental property of placing any single peak at the correct frequency. The Equations (3) relating $\hat{V}$ and $\hat{\hat{V}}$ (or $\hat{S}$ and $\hat{\hat{S}}$) no longer hold, as different methods of applying the taper have been used. The sidelobes of $\hat{\hat{V}}$ seem to have been reduced considerably, and the Vaníček method seems to show up quite well.

Acknowledgements

Mr G. Holland first suggested that this project should be undertaken, and commented upon the various earlier versions of the paper. The method of analysis used here is an extension of an approach used by Dr G. Godin in an unpublished critique of the Vanicek method, and we thank Dr Godin for his critical review of this article. We would like to thank Mr P. Richards who did a considerable amount of the computer work for this study.

References

Crow, E., Davis, F., and Maxfield, M.: 1960, *Statistics Manual*, Dover, New York.
Enochson, L. and Otnes, R.: 1968, *Programming and Analysis for Digital Time Series Data*, Shock and Vibration Center, Naval Research Laboratory, Washington, D.C.
Henry, R. and Graefe, P.: 1971, *Zero Padding as a Means of Improving Definition of Computed Spectra*, Marine Sciences Branch Manuscript Report Series No. 20, Environment Canada, Ottawa.
Vaníček, P.: 1971, *Astrophys. Space Sci.* **12**, 10–33.