Spectral fidelity of gappy data

By FERDINAND BAER and JOSEPH J. TRIBBIA, *Department of Atmospheric and Oceanic Science, The University of Michigan, Ann Arbor, USA*

(Manuscript received March 17, 1975)

ABSTRACT

A one-dimensional periodic domain is chosen over which four physical variables are presumed to be known exactly at 72 equally spaced points. Nineteen experiments are then designed with differing numbers of observing points thus yielding variable gaps in the data. Utilizing trigonometric functions as a basis, Gram-Schmidt functions are generated with orthogonal properties over the non-equally spaced point domain. The quality of the Fourier coefficients for each allowed scale is assessed for the different experiments in terms of the true Fourier coefficients. Results indicate that accuracy in scale coefficients deteriorates rapidly for scales less than twice the maximum gap in the domain, regardless of point density elsewhere in the domain. Furthermore, for the addition of observations in a gappy region, the most spectral fidelity will be achieved by minimizing the largest resulting gaps.

1. Introduction

Traditionally atmospheric parameters have been sampled regularly in time but rather erratically in space. Such spatial irregularity in data has always been burdensome to analysts and no less so to current computer analysts. Indeed since demand for initial conditions has increased substantially with numerical forecasting based on models requiring regular grid spacing, devices such as smoothing operators have been applied to data sets. These operators convert data from their station locations to a regular net of coordinate locations convenient to prediction models. Smoothing is however not independent of prejudice and consequently imposes scale dependent biases which may or may not exist in the observed data set. Such biases may mask the true scale dependence of data or delude the observer into accepting scale information which is not representative of the original observations. Spatial irregularity of data and its significance has not eluded the critical eye of the meteorological community as may be observed from the studies of Jones (1972), Julian et al. (1970), Yang & Shapiro (1973) and Leory & Thompson (1973). If data are converted to scale dependence by some suitable transformation—an example would be Fourier transforms on a latitude circle—an analysis may determine the quality to be expected in each scale. The results of such an endeavor based on large atmospheric scales was presented by Julian et al. (1970) without recourse to smoothing.

Continual improvement of numerical prediction models also extends the demands on data for verification purposes as well as initial conditions. Additionally the need for closure approximations based on scale considerations requires statistical information from observations which may substantiate theoretical conclusions. Despite the need for accurate short scale observational information in both of the above, these are the scales which are least reliably described by irregularly spaced observing stations.

To understand the distribution of atmospheric variables with scale, a frequently used device is to convert the data from space dependence to wave amplitude by a spectral transformation. Formally one may make this transformation such that the number of scales described is half the number of data points and the scales are commonly taken sequentially in size beginning with the largest, with the hope of minimizing aliasing errors. Results of such spectral analyses, principally of smoothed data, by Horn & Bry-

son (1963), Wiin-Nielsen (1967), Julian et al. (1970), Baer (1972) and Chiu (1973) have lead to comparisons between observed shorter scale energy distributions and theoretical expectations (Kraichnan, 1967; Leith, 1971; Charney, 1971). Because of the non-regular data observations, and the consequent degradation of quality in the shorter scales, some question remains concerning the significance which can be placed on the observed spectra. We thus require further information on the influence of erratically spaced data (gappy data) on spectral quality.

To determine the spectral quality which one might anticipate from erratically spaced data utilized *without* smoothing, we have performed the following experiment. Let us presume a set of data known at equally spaced points along a latitude circle which we shall convert to spectral form by a Fourier transformation, describing exactly the first allowed N waves. The choice of a one-dimensional experiment was made for convenience and ease of description with the extension to two-dimensions being apparent, and the assumption of exactly known data was made to bypass the difficult question of aliasing. We now define a number of subsets of this given data set with the subsets having observations at some but not all of the points defined by the original set. In this way we create unequal gaps in the observation set which may reflect true gaps in atmospheric observations such as ocean areas. The data utilized at the specified points of the subsets are identical to the data on the original set. Since the subset points are no longer equally spaced, the Fourier transform becomes less straightforward. We have chosen to generate Gram-Schmidt functions based on the Fourier series to simplify this transformation, a process used successfully by Fougere (1963) and Lee & MacDonald (1963), and applied to atmospheric data analyses by Dixon et al. (1972). Given the transforms of the subsets, we may compare the Fourier coefficients of the original complete set (considered accurate) to those of the subsets and determine the effective quality of the various scales in terms of the gappiness in the data. To avoid a data bias, we have applied the experiment to the fields of wind, temperature and height. The procedure outlined above is but one method of arriving at spectral quality of gappy data, and it is clearly not the only nor necessarily the best one. It does, however, elucidate the problem and is furthermore consistent with the general practice of generating spectra from grid data. Since grid data is smoothed from unequally distributed reporting stations, the quality of the spectra taken from grid data cannot be assessed, and our experiments will attempt to give an indication of the possible errors derived from smoothing, using a direct expansion approach. We do not suggest how improved spectra might be generated although our results will clearly indicate the need for such a development. One could, for example, incorporate statistical information and thus effectively expand the original data set. Alternately one could utilize spectra generated from a covariance function (see for example Yang, 1974). Since the theory for covariance spectra is based on stationarity of time series, the utilization of this assumption in the space domain must be justified. This should be done and the results of such an analysis compared to our results. Were the covariance spectra to yield superior short scale fidelity, additional and worthwhile information would thus be made available. The experiments we have outlined, however, should expose the spectral fidelity problem in a self-consistent fashion.

2. Methodology

Let us consider a one-dimensional periodic domain such as a latitude circle on which data observations are available. These observations are discrete and may be taken uniformly or non-uniformly. We will assume that N observations are available and when $N = N_{\max}$ we will require the observations to be exact and equally spaced. Any N observations will be a subset of $N_{\max}$ and the values assigned to the chosen (N) locations will be identical to those from the corresponding complete ($N_{\max}$) set. Since we wish to discuss the scale characteristics of the data, and since our domain is periodic, a convenient procedure is to expand the data in a Fourier series; the wave number will then be uniquely related to scale and we can discuss the Fourier amplitudes with regard to scale. We have selected our basis functions as follows:

$$G_i(\lambda_m) = \sin\left[\frac{i}{2}\right]\lambda_m + (-)^{i+\delta_{1i}}\cos\left[\frac{i}{2}\right]\lambda_m \qquad (1)$$

where the bracket notation denotes integral value, $0 \leqslant \lambda_m \leqslant 2\pi$ with λ the coordinate along our domain and $1 \leqslant m \leqslant N$, so that λ_m represents a point in the subset of N points. Expanding any data sample in terms of these basis functions would yield at most $N/2$ scales.

Unfortunately, unless $N = N_{\max}$ (equal spacing), the functions G_i are not simply orthogonal over the domain. Conventional procedure would suggest smoothing the data to a set of equally spaced points. In order to avoid arbitrary smoothing, we utilize the Gram-Schmidt technique (see Fougere, 1963) to generate orthogonal functions over the non-regular domain of N points. These functions will be orthogonal and normal when defined in terms of the basis functions as

$$H_j \equiv b_j \left\{ G_j - \sum_{k=1}^{j-1} B_{kj} H_k \right\} \quad 2 \leqslant j \leqslant M \tag{2}$$

subject to the constraint that

$$\sum_m H_j(\lambda_m) H_i(\lambda_m) = \delta_{ij} \tag{3}$$

Utilization of (3) in (2) will show the coefficients B_{ij} and b_j to be given in terms of known functions as

$$B_{ij} = \sum_m H_j(\lambda_m) H_i(\lambda_m) \quad 1 \leqslant i < j$$

$$b_i = \left[\sum_m G_j^2 - \sum_{k=1}^{j-1} (B_{kj})^2 \right]^{1/2} \tag{4}$$

which may be evaluated recursively.

The scaling characteristics of the H's and G's are similar, although not identical, and the Gram-Schmidt functions may be written in terms of the basis functions by the series

$$H_j(\lambda_m) = \sum_{i=1}^{j} A_{ij} G_i(\lambda_m) \tag{5}$$

so that the H's have at most as many zeros as the corresponding G's. If we define $B_{jj} \equiv 1/b_j$, the A's may be developed from the B's by the formula

$$\sum_{i=k}^{j} A_{ik} B_{ji} = \delta_{kj} \quad 1 \leqslant k \leqslant j \tag{6}$$

It is evident from (5) that the characteristic amplitudes of the H functions need not be unity as they are for our basis functions. Indeed a characteristic amplitude of any H may indicate the quality of the Fourier coefficient, since for equal spacing the H's and G's coincide and both amplitudes are unity. We define the characteristic amplitude of any H function (P_j) from the integral of its variance over the domain; thus,

$$P_j \equiv \frac{1}{2\pi} \int H_j^2 \, d\lambda = \sum_{i=1}^{j} A_{ij}^2 \tag{7}$$

These amplitudes will differ for each subset of N points and we shall discuss their scale distributions subsequently.

Having now established the orthogonal set of functions, we may expand the given data at any point in the series

$$F(\lambda_m) = \sum_{j=1}^{N} C_j H_j(\lambda_m) \quad 1 \leqslant m \leqslant N$$

$$C_j = \sum_m F(\lambda_m) H_j(\lambda_m) \tag{8}$$

In terms of the Fourier coefficients,

$$F(\lambda_m) = \sum_{i=1}^{N} A_{iN} G_i(\lambda_m)$$

$$A_{iN} = \sum_{j=i}^{N} C_j A_{ij} \tag{9}$$

Note that the best fit for any subset of N points is given by (8) for the first M ($M \leqslant N$) coefficients C_i. This is clearly not the case for the Fourier series since if we truncate (8) at $M < N$, the coefficients A_{iM} will be seen to differ from those for A_{iN}.

Since we assume the exact data to be given for $N = N_{\max}$, we may assess the quality of any Fourier coefficient by measuring its value relative to the coefficient $A_{iN_{\max}}$. However, since the Fourier coefficients may also change for different truncations within any N subset if we choose the truncation $M < N$, we shall measure the quality in any Fourier coefficient by the ratio,

$$f_{iN}(M) \equiv \frac{A_{iM}}{A_{iN_{\max}}} \quad M \leqslant N \tag{10}$$

We shall see subsequently that choosing $M = N$ does not always provide the "best" coefficient unless $N = N_{\max}$.

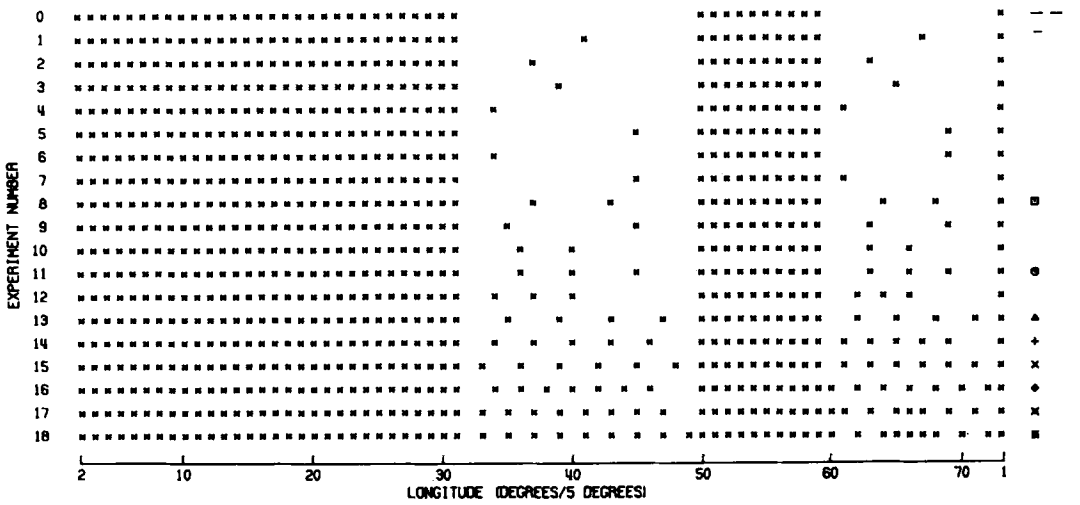

Fig. 1. Nineteen experiments utilized in this study. Asterisks denote observation points and minimum distance between points is 5° longitude. A maximum of 72 points comprise the periodic domain.

Having established the series expansion (8) from the given data set, it is possible to re-evaluate the functional value F at any point in the domain for a truncated series terminating at $k \leqslant N \leqslant N_{\max}$. This value may be compared to the true value given when $k = N_{\max}$. We will determine errors from the "true" values in the RMS sense as a function of wave truncation as follows. We first define $F_t(\lambda_m)$ to be the "true" value at each point λ_m. We next define $F_k(\lambda_m)$ to be the value at each point of the set $N_{\max}$ calculated from the subset N and truncating the series at $k \leqslant N$. The root mean square error which we find over the domain due to gappy data ($N < N_{\max}$) and truncating at wave $k \leqslant N$ will be given by the error relation

$$E_t(k, N) \equiv \left[\frac{1}{N_{\max}} \sum_{m=1}^{N_{\max}} (F_t(\lambda_m) - F_k(\lambda_m))^2\right]^{1/2} \qquad (11)$$

It may be expected that the errors in approximation would be different if tested over those points in the subset of N points which are in the total set as compared to those which are not in the subset. These RMS errors are defined as E_g and E_b respectively and may be calculated from the following formulae:

$$E_g(k, N) \equiv \left(\frac{1}{N} \sum_{i=k+1}^{N} C_j^2\right)^{1/2}$$

$$E_b(k, N) \equiv \left[\frac{N_{\max}}{N_{\max} - N}\left(E_t^2 - \frac{N}{N_{\max}} E_g^2\right)\right]^{1/2} \qquad (12)$$

These error measures should give us some insight into the effect of gappy data on different scales and the quality which might be expected from these scales, as represented by our experiments.

3. Experiments

The experiments we have chosen are designed to have some relevance to the atmospheric observation network. We begin with the assumption that data is available on an equally spaced mesh with $N_{\max} = 72$ points and that the domain is periodic. If the domain represents a latitude circle, we would find that $\lambda_{m+1} - \lambda_m = 5°$ longitude for the complete set of $N_{\max}$ points. Subsequently for subsets with gaps in the observations, $\Delta\lambda_m$ may be some multiple of 5°; i.e., $\lambda_{m+1} - \lambda_m = 5S$ where $S \geqslant 1$ and is an integer. We have generated 19 experiments or subsets, each with $N < N_{\max}$ and for 10 uniquely different values of N. Some of the experiments thus have the same number of observation points (N) but their distribution is different. The reason we have chosen these latter experiments is to determine how sensitive the quality of scale elements is to the data distribution given a fixed number of observation points.

The experiments are depicted graphically on Fig. 1 wherein each asterisk denotes an ob-

serving point along the domain $0 \leqslant \lambda \leqslant 2\pi$. The first experiment might correspond to the mid-latitude situation of a latitude circle with reporting stations dense over two continents and no information over two oceanic regions. In the experiment this is described by 31 contiguous reporting stations with $S = 1$, a gap of 18 missing stations, followed by 10 contiguous reporting stations and finally another gap with 13 missing stations. Subsequent experiments as described by Fig. 1 are developed by the gradual addition of reporting stations in the gapped domains of the next less dense experiment. Note that experiments 1–7 have a common value of N but different distributions of points, as do experiments 8–10, and 11–12. Analysis over these subsets should indicate the effect of data distribution for an equal number of reporting stations.

The behavior of the Gram-Schmidt functions generated from each of these experiments may be seen from several represented on Fig. 2. We have chosen to describe the Gram-Schmidt functions for index numbers 2, 7, 12, 17 and 22 (see (2) for indexing) over the domain $0 \leqslant \lambda \leqslant 2\pi$. The Gram-Schmidt functions describe periodic behavior similar to the basis functions but their amplitude structure varies significantly. For the lowest scale graphed, index no. 2, the deviation from simple sinusoidal form varies with experiment number, showing large amplitude fluctuations for those experiments with fewer reporting stations; note that only those experiments with different N values have been plotted. Furthermore the largest deviation occurs in the regions of maximum gap. These observations are consistent for all five scales described with the amplitude fluctuations becoming more severe for smaller scales. Indeed the departures from the simple basis functions become overwhelming, as may be seen from the ordinate scaling, and the experiments with smallest N clearly show the largest amplitude oscillations, again in the region of missing data (maximum gap).

One might anticipate representational difficulties with functions containing large amplitude fluctuations. Such fluctuations may be characterized by the characteristic amplitudes defined in (7). If the characteristic amplitudes exceed unity significantly, a satisfactory expansion would require correspondingly small magnitudes in the expansion variables derived from

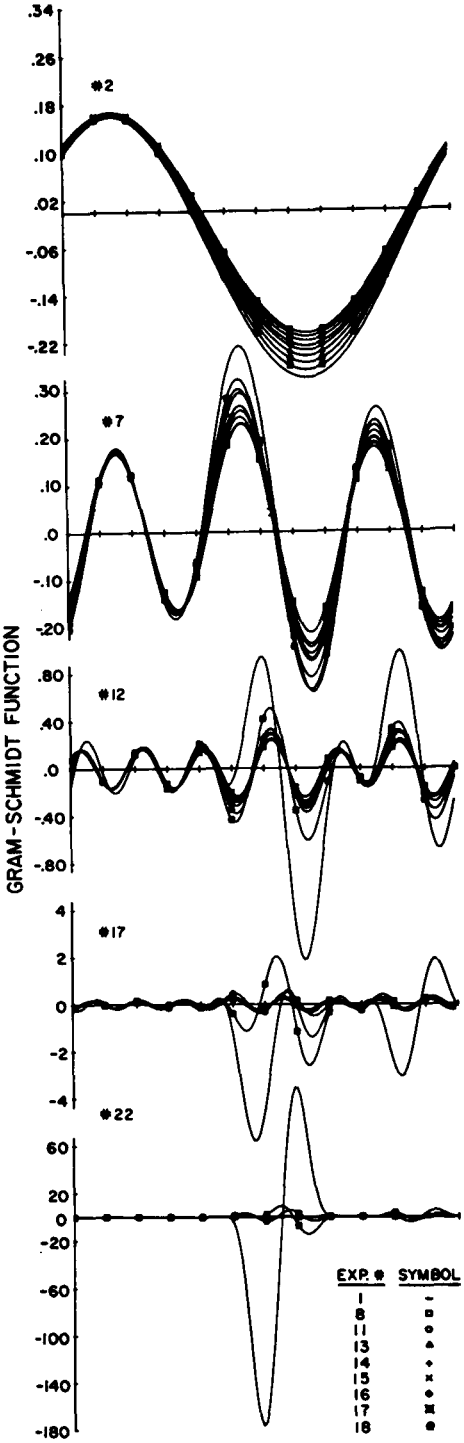

Fig. 2. Gram-Schmidt functions for five different indices plotted on the periodic domain, $0 \leqslant \lambda \leqslant 2\pi$.

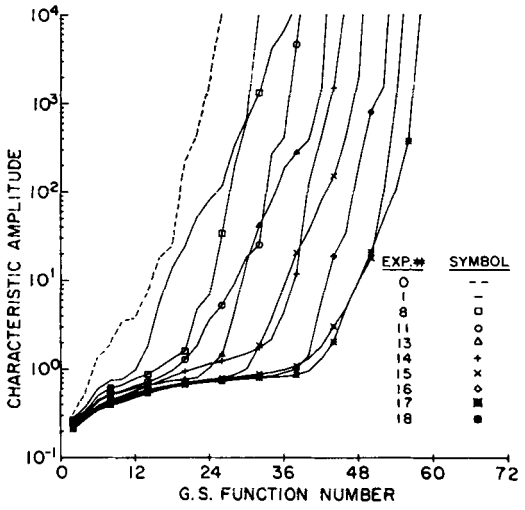

Fig. 3. Characteristic amplitudes of the Gram-Schmidt functions (defined in text) in terms of the function index and for all experiments with different total number of observing points.

the data according to (8) or (9); these expansion coefficients C_j or A_{iN} will be discussed in the sequel. We have described on Fig. 3 the characteristic amplitudes (see eq. 7) for each Gram-Schmidt function and for those experiments which have different N values. It should be noted that the Gram-Schmidt function numbers increase by two for each additional scale element represented in order to account for phase. Fig. 3 shows that to index 12, all characteristic amplitudes are of order unity or less. For larger indices (smaller scales) the amplitudes increase dramatically, an observation already anticipated from Fig. 2. This amplitude growth is strongly dependent on experiment number, with those experiments having most observing stations representing more scales with reasonable characteristic amplitude.

Fig. 3 gives us advanced indication that serious scale errors are to be expected in the shorter scales for data sets with large gaps. If we wish to have adequate representation from our Gram-Schmidt functions, we might wish to truncate our series at those scales for which the characteristic amplitudes do not substantially exceed unity. Thus the shortest scales would be omitted as nonrepresentative although they could be theoretically included. We shall see the implications of this observation subsequently when we evaluate real data samples.

4. Data

To test the spectral quality of expansions based on (8) for the different experiments shown on Fig. 1, we have selected the four observed fields which include temperature, height, u and v wind components. This data, made available from NCAR, is taken from the 45° N latitude circle at 500 mb every 24 hours including January and February of 1971. The observed or analysed values, originating from NMC, are given at each of the $N_{\max}$ points, thus providing information for each variable at every five degrees of longitude. Although the original data may have been smoothed to these points, we assume that they are accurate for our study and proceed on this assumption. Thus for any experiment with $N < N_{\max}$, we choose the values from the original set as given above, neglecting of course those points which are not included in the subset N. Calculations are made with this data for each time period and the resulting quantities, such as $f_{iN}(M)$ and $E_t(k, N)$, are averaged over all time periods.

5. Analysis

We have already indicated our expectations of spectral quality based on the characteristic amplitudes described by Fig. 3. Application of our expansions given by (8) or (9) to the data sets discussed in the previous section will show how scale dependent errors develop.

Let us first consider the root mean square error from the true data given at all points of the domain by the data determined from some experiment with N observations, utilizing only the first $k \leq N$ components of the series expansion. This measure (E_t) is defined by eq. (11) and its development for the two experiments 0 and 18 (see Fig. 1 for experiment configurations) as a function of the truncation index k is described on Fig. 4 by the solid curves. The physical variable under consideration here is temperature; however, we shall note subsequently that the behavior characteristics for all physical variables considered (H, T, U, V) are remarkably similar. Our first observation from this figure is the well defined minimum in the error corresponding to a given wave index, the minimum index differing substantially between the two experiments. After this mini-

mum, which also occurs at a different magnitude for the two experiments, the error rises rapidly suggesting that the best representation for a given experiment should occur when the series is truncated at the scale index (k) of minimum error. For those experiments described by Fig. 4, substantially more waves would be included for experiment 18 than for experiment 0. We also note that the rise in error with scale index is similar to the rise in characteristic amplitude as seen from Fig. 3.

The error currently under discussion is derived from an evaluation over all points in the domain (N_{max}). However, for any experiment with N points, there are $N - N_{max}$ points on which no information is available but for which values are derived from the series expansion; the expansion procedure does not require these points to be fit well so long as the N points on which data are available are properly fit. How well all points included in the Nth experiment are fit is described by the root-mean-square error E_g while the fit to the points not included in the experiment are described by E_b; both measures are defined by eq. (12). These errors are included in Fig. 4 as the dashed curves and are appropriately identified with the number in brackets denoting experiment number. We

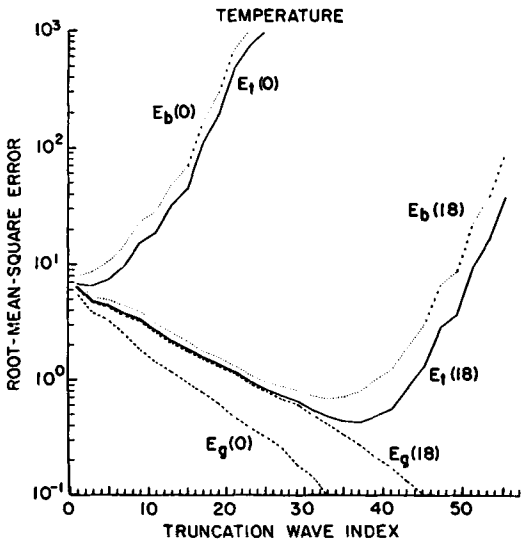

Fig. 4. RMS error of temperature data for experiments 0 and 18 as a function of truncation wave index. Included are errors over the total domain, over the domain of observing points and over the domain of points with no observations.

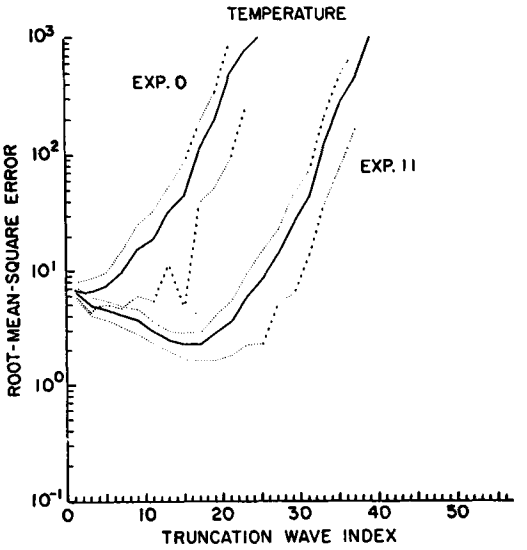

Fig. 5. Variance from the time mean of RMS errors in temperature for experiments 0 and 11 in terms of the truncation wave index.

see immediately, as expected from theory, that E_g approaches zero as $k \to N$. We see also that following a minimum in E_b, this error grows rapidly with increasing k and dominates the error in the total region (E_t). Therefore both E_b and E_t follow one another closely. Although no information is available at the points utilized in calculating E_b, out to the scale of minimum error the representation at these points is satisfactory. The significance of this result will become apparent when we discuss the errors in the Fourier coefficients.

As indicated previously, the calculations presented have been performed at each time period of the data set but the results have been averaged in time. To assure ourselves that the time mean values are representative of the individual values we have prepared Fig. 5 which includes the distribution of E_t with truncation index k for the two experiments 0 and 11 utilizing temperature data. Since comparable results were found for all experiments and physical variables, only these examples are depicted. Together with the solid curves of E_t describing time averages we have drawn the corresponding curves (dashed) which include the time variance. Although the figure shows significant variance, the basic behavior of E_t is unaltered from one time period to another

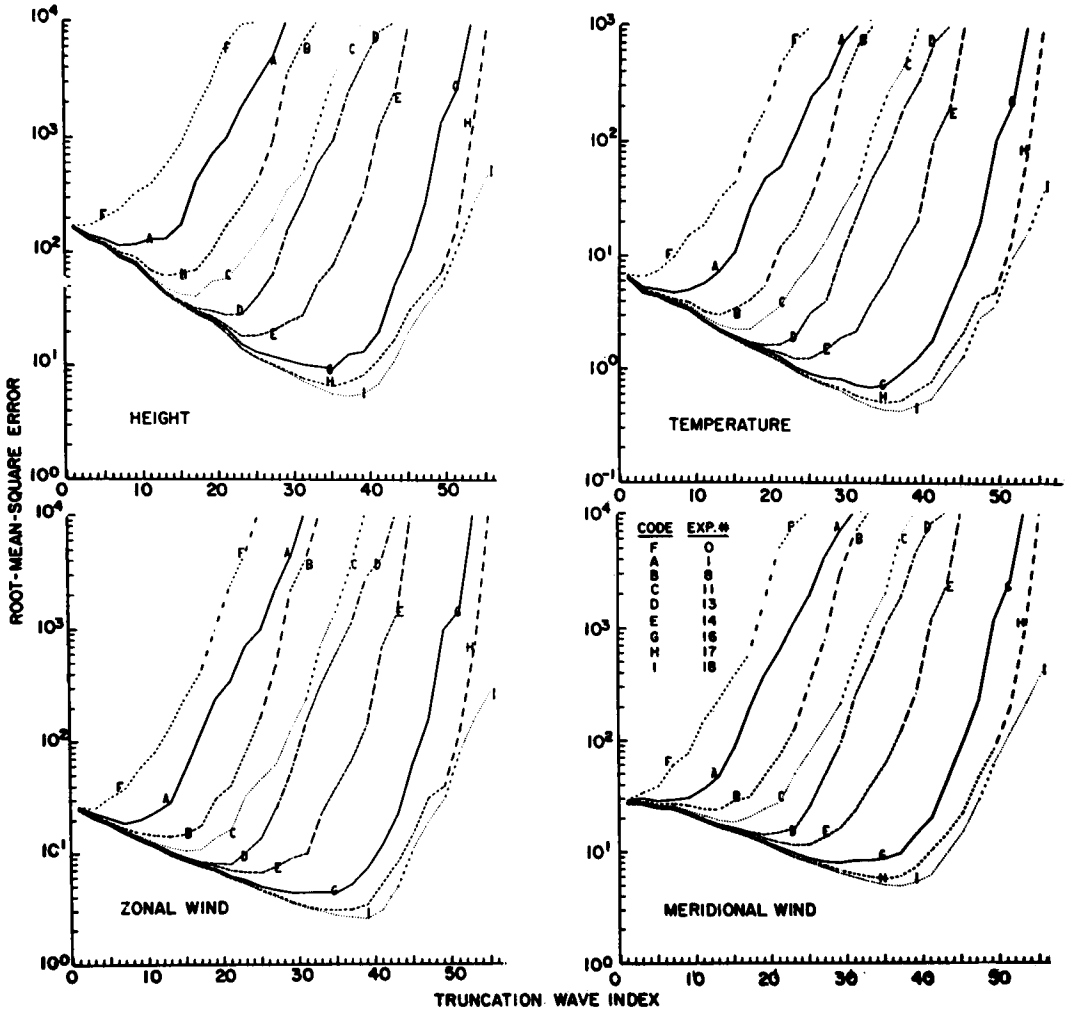

Fig. 6. RMS errors in terms of truncation wave index for experiments 0, 1, 8, 11, 14, 16, 17, 18. Data described includes temperature, height and u, v wind components.

and our conclusions may be discussed legitimately from the time-mean values.

The examples of E_t seen in Figs. 4 and 5 suggest a systematic tendency of the error curves to have a minimum and that this minimum moves to higher truncation index with increasing experiment number, the higher experiment number indicating more data resolution. Fig. 6 shows the error curves of E_t for all experiments with different N values—i.e., with differing numbers of observation points—omitting only experiment 15 because of its similarity to experiment 14. For completeness we include the curves for all four physical variables. Although the amplitudes of the errors as reflected by the ordinate scales differ for each physical variable, the shape of the curves are remarkably similar, leading to the conclusion that the quality of any expansion in terms of wave number does not depend on the physical variable to be represented, but on the number and distribution of the observing points. The close similarity between the curves on Fig. 6 and those on Fig. 3 (characteristic amplitudes) corroborates this conclusion. For any one data set we see how the wavelength of minimum error does

indeed increase systematically with increasing resolution together with a corresponding decrease in the error magnitude. Since the different experiments are gap dependent, and since the errors they generate are clearly scale dependent (as evidenced from Fig. 6), we must conclude that the gap characteristics of the different experiments are distinctly scale dependent.

To establish the correspondence between data gaps and spectral (scale) resolution as implied by Fig. 6, we have prepared Table 1. Noting from Fig. 1 that each experiment has a maximum interval of length $S_{max}\Delta\lambda$ with no observations—the integer notation S was assigned to intervals in Section 3—we have listed this interval in the table following its corresponding experiment number. We next calculate the number of observing points required for each experiment if the points were equally spaced with interval $S_{max}\Delta\lambda$, and also list this result in the table. Noting the orthogonality of finite Fourier series over an equally spaced domain of points, the number of derivable wave components (twice the number of waves) is equal to the number of points. Finally we have determined the wave index of minimum error for each experiment and each physical variable,

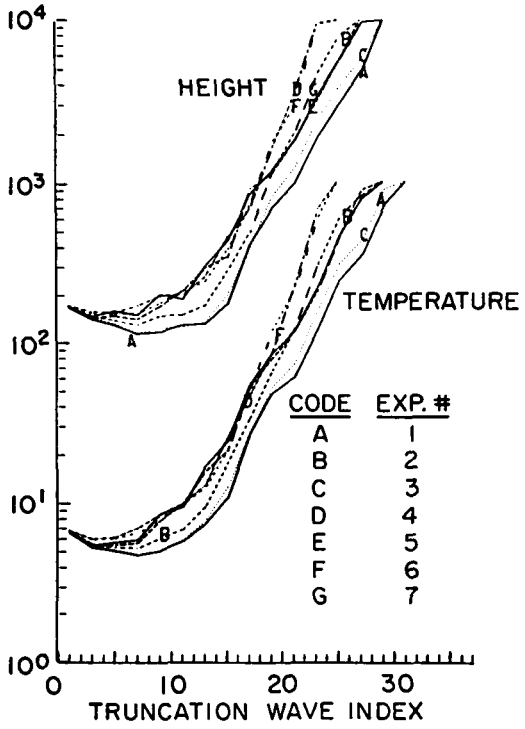

Fig. 7. RMS errors in experiments 1–7 for both temperature and height. These experiments have the same number of points with different distributions.

Table 1. *Root-mean-square error statistics of height, temperature and wind components*

Exp. no.	Max. increment (S_{max})	Min. no. points equally spaced	Wave index of min. error			
			H	*T*	*U*	*V*
0	19	3	3	3	3	1
1	10	7	7	7	7	7
2	13	5	7	7	7	5
3	11	6	7	7	7	7
4	16	4	3–4	4	3	5
5	14	5	3–4	4	3	5
6	16	4	3–4	4	3	5
7	14	5	3–4	4	3	5
8	7	10	13	13	13	12
9	10	7	11	11	9	11
10	10	7	11	11	9	11
11	5	14	17	17	17	15
12	10	7	13	12	7	10
13	4	18	22	20	21	19
14	4	18	23	23	23	25
15	3	24	26	27	27	27
16	4	18	35	34	35	33
17	3	24	35	35	36	35
18	2	36	37	37	39	38

some of which are apparent on Fig. 6, and listed them in Table 1. Perhaps not too surprisingly, the wave indices of minimum error correspond closely and with few exceptions to the minimum number of equally spaced points based on maximum gaps. The implications of this result are apparent. Little spectral quality is derived from data-rich regions (small gaps) if there are large gaps in the domain. Thus we see that experiment 0, which has a gap of 19 basic intervals, gives adequate results only for the first wave, and the addition of information from shorter waves generates large errors. In contrast, experiment 18 shows high quality results with the inclusion of the first 20 waves, but the addition of further wave information causes rapid error growth. Several experiments are not consistent with this observation (12, 16 and 17 as examples) and suggest that more than the minimum number of scales, based on the largest gap, may be represented with some fidelity for unequally spaced data sets. Never-

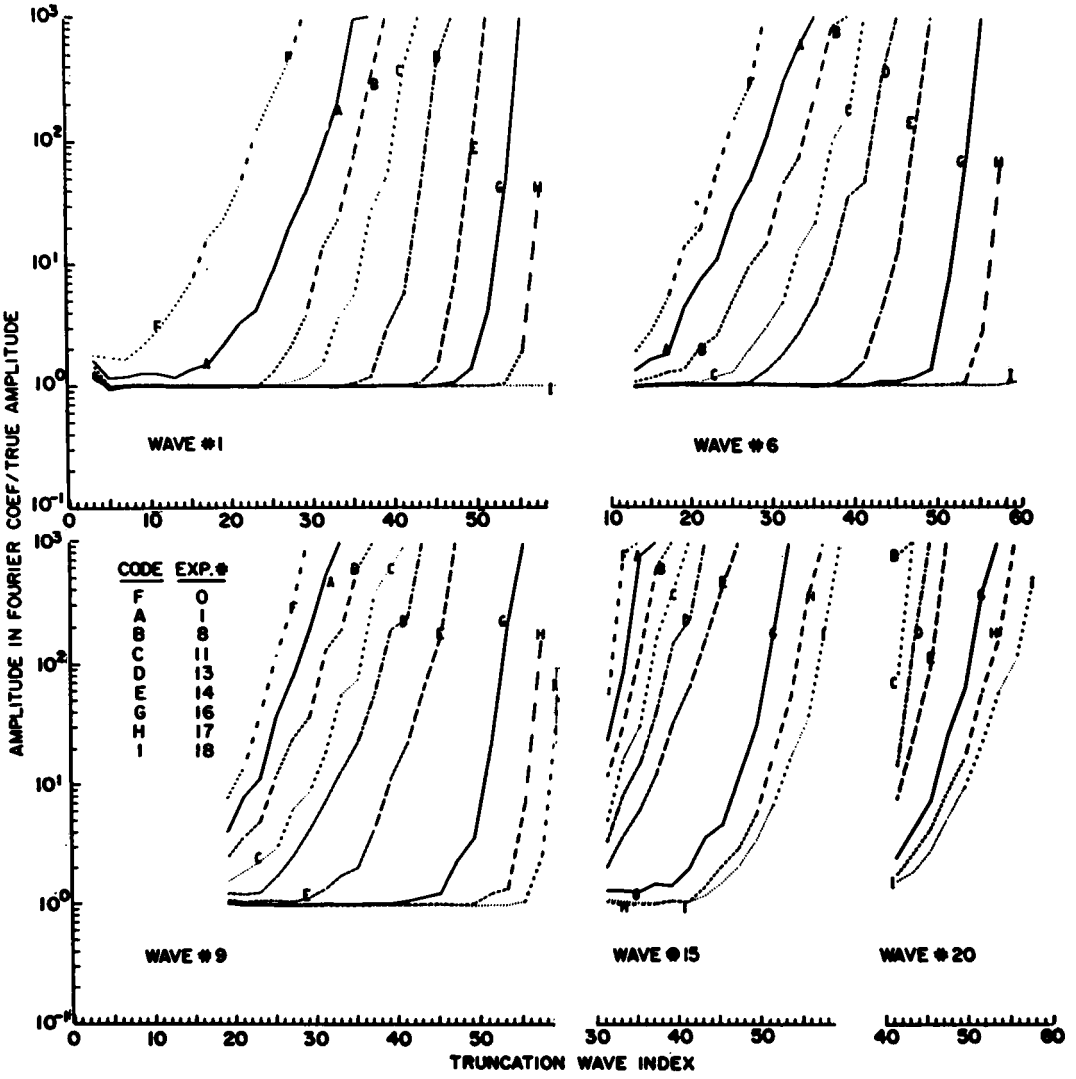

Fig. 8. Fourier coefficients as a fraction of their true value for experiments 0, 1, 8, 11, 14, 16, 17, 18 in terms of the truncation wave index for the 500 mb height field. Five different waves are described.

theless we would not be far from accurate to state that the spectral fidelity of a one-dimensional data set is determined to a large extent by the largest gap in the domain; i.e., waves with wave length substantially less than twice the largest gap will more than likely have unsatisfactory amplitude representation. We must note that this conclusion is based on the direct inversion of our data sets via an orthogonal transformation to spectral coefficients and may not be consequent for other spectra generating techniques.

That the spectral resolution possible from a given subset of data depends not so much on the number of observing points (N) as on their distribution, an observation already elucidated above, is emphasized in Fig. 7 where we show the E_t curves for experiments 1–7, all of which have the same number of data observations (N) but different distributions (see Fig. 1). Experiment 1 has the minimum error and largest minimum wave index for both temperature and height (and indeed for U and V wind components) whereas experiments 4 and 6 have the

largest errors. A glance at Table 1 will show that of the experiments graphed, experiment 1 has the smallest maximum gap whereas experiments 4 and 6 have the largest. Thus, should we in our experiments desire maximum spectral resolution from the addition of one observation in a gap, that observation should be made to minimize the gap; i.e. it should be centered in the open region.

As noted in Section 2, eq. (9), the Fourier coefficients (A_{IN}) developed for each experiment depend not only on the N value of the experiment, but also on the truncation of the Gram-Schmidt functions if we choose to determine those coefficients from less than all the allowed Gram-Schmidt coefficients (C_j). Nevertheless the exact value of each Fourier coefficient is given for $N = N_{max}$. Thus we develop a measure of the quality of the Fourier coefficient for a given experiment and given truncation to its exact value in the parameter $f_{IN}(M)$ defined by eq. (10). Fig. 8 describes this parameter for selected wave *numbers* for all experiments with different N values (again excluding experiment 15) in terms of the truncation wave *index*; the applicable data set, representative of all the data sets is the 500 mb height field. So long as the parameter, f, remains near unity, the Fourier coefficient is satisfactorily represented. We have chosen to represent wave numbers in this figure to avoid phase considerations. Let us consider the curves for wave number one. Here we see that for experiment 0, utilizing information from Gram-Schmidt coefficients with indices in excess of 7 or 8 leads to substantial errors in the Fourier coefficients (there are two coefficients per wave). In contrast, utilizing all Gram-Schmidt coefficients for experiment 18 leads to effectively no errors in the Fourier coefficients. The other experiments all show some truncation limit beyond which the Fourier coefficients are in serious error. The curves for waves 6, 9, 15 and 20 show similar results. However, in these waves we already see that some experiments cannot yield adequate representation for the Fourier coefficients. As an example, experiments 0, 1, and 8 cannot give an adequate representation for the Fourier coefficients of waves greater than or equal to nine, and only the first few Gram-Schmidt coefficients will yield satisfactory Fourier coefficients for wave nine with other experiments.

These results are consistent with and might well have been anticipated from Fig. 6, since large amplitude errors in Fourier coefficients would lead to correspondingly large root mean square errors over the domain. The errors in Fourier coefficients with Gram-Schmidt index truncation is considerably more sensitive than might be anticipated from the characteristic amplitudes described in Fig. 3, although we may utilize the characteristic amplitudes as a lower bound on the errors in the Fourier coefficients. A close scrutiny of Figs. 3 and 8 will show substantial errors in Fourier coefficients (by factors of 3 or more) for characteristic amplitudes of order unity. We are again, however, confronted with the striking observation that the quality of any Fourier coefficient is strongly related to the largest gap in the observation set.

6. Conclusion

Scale characteristics of atmospheric data are frequently derived from data smoothed to an equally spaced grid and the spectral coefficients are determined by a direct inversion of this data. The smoothing masks the irregularity of the observing stations and the quality of the spectral coefficients as a function of scale is difficult to assess. To affect this assessment, we have expanded some atmospheric parameters on a one-dimensional domain for a number of different station distributions. Beginning with the assumption of exact information on a uniform grid we generated the "exact" Fourier coefficients with which to compare the coefficients derived from the truncated data sets. While interpreting the results of our experiments noted below, the reader should bear in mind that they stem from the determination of spectra by direct inversion, the process commonly used by meteorologists to draw conclusions concerning the scale dependence of atmospheric parameters. That alternate methods exist for establishing spectra is not disputed nor do we lay claim to utilizing the "best" method. Because of its frequent usage, we feel justified in the procedure we have adopted and our results are self-consistent.

The quality of the Fourier coefficient derived for a given distribution of observing stations depends strongly on the largest gap in the domain. Wave components with wave lengths

greater than or equal to twice this maximum gap are well represented. Shorter waves show significant errors from their "true" values.

One must take care in interpreting amplitudes representing scales smaller than the largest gap interval. Moreover, this result is not significantly modified by high data density in subregions of the total domain, provided the maximum gap is not reduced.

The availability of high data density in local regions may lead to serious misinterpretations in the spectral amplitudes of shorter scales. Should one be led to interpolate data into regions of sparse data density, scale information can be determined for wavelengths for smaller than the maximum gap. Such information could well be incorrect. Its fidelity would be directly related to the procedures used for data interpolation. If the interpolation is based on physical premises (see for example Thompson, 1961) meaningful results could be achieved. However, most physical interpolation is based on scale characteristics of the system larger than those for which the interpolation is desired.

A detailed analysis of spectral information in data-rich regions, now readily available from satellites, may yield insight into the spectral structure of scales smaller than the largest gap over the hemisphere. Should local regions describe spectral structures on shorter scales which are quasi-invariant in both space and time, such information could legitimately be interpolated into regions with large gaps. The subsequent scale analysis (Fourier transform in one dimension) would lead to acceptable spectra which could ultimately be used for model closure conditions or other possible applications.

Concurrent with finding suitable methods for data interpolation, alternate procedures for establishing spectra should also be tested. These include, as indicated earlier, the determination of spectra from covariance functions if the concept of stationarity in time series can be translated into spatial terms. Barring dramatic changes in spectral quality as a function of scale by alternate methodologies, if one has the option of setting up a station network and one is concerned to obtain the most fidelity in the corresponding scale spectra, our results indicate that observations should be taken at equally separated points along the domain. Alternately, if one wishes to fill a large gap with one or more observations, the best spectral resolution will be obtained if these observations are equally spaced in the gappy region.

Finally we should note that the results discussed above relating to the quality of scale coefficients depend primarily on the observing station distribution and very weakly on the nature of the data itself. Thus one should be able to establish the spectral fidelity of gappy data simply from the locations of observations in the domain.

Acknowledgements

The research reported herein has been supported by Grant GA-36409 from the National Science Foundation, Atmospheric Series Section, to the University of Michigan. Computing time was made available by the Computing Facility of the National Center for Atmospheric Research which is supported by the National Science Foundation.

REFERENCES

Baer, F. 1972. An alternate scale representation of atmospheric energy spectra. *J. Atmos. Sci. 29*, 649–664.

Charney, J. G. 1971. Geostrophic turbulence. *J. Atmos. Sci. 28*, 1087–1095.

Chiu, Wan-Cheng. 1973. On the atmospheric kinetic energy spectrum and its estimation at some selected stations. *J. Atmos. Sci. 30*, 377–391.

Dixon, R., Spackman, E. A., Jones, I. & Francis, A. 1972. The global analysis of meteorological data using orthogonal polynomial base functions. *J. Atmos. Sci. 29*, 609–622.

Fougere, P. F. 1963. Spherical harmonic analysis. *J. Geophys. Res. 68*, 1131–1139.

Horn, L. H. & Bryson, R. A. 1963. An analysis of the geostrophic kinetic energy spectrum of large-scale atmospheric turbulence. *J. Geophys. Res. 68*, 1059–1064.

Jones, R. H. 1972. Aliasing with unequally spaced observations. *J. Appl. Meteor. 11*, 245–254.

Julian, P. R., Washington, W. M., Hembree, L. & Ridley, C. 1970. On the spectral distribution of large-scale atmospheric kinetic energy. *J. Atmos. Sci. 27*, 376–387.

Kraichnan, R. H. 1967. Inertial ranges in two-dimensional turbulence. *Phys. Fluids 10*, 1417–1423.

Lee, W. H. K. & MacDonald, G. J. F. 1963. The

global variation of terrestrial heat flow. *J. Geophys. Res. 68*, 6481–6492.

Leith, C. E. 1971. Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci. 28*, 145–161.

Leory, C. & Thompson, R. O. R. Y. 1973. Shortcomings of an objective analysis scheme. *J. Appl. Meteor. 12*, 589–594.

Thompson, P. D. 1961. A dynamical method of analysing meteorological data. *Tellus 13*, 334–349.

Wiin-Nielsen, A. C. 1967. On the annual variation and spectral distribution of atmospheric energy. *Tellus 19*, 540–559.

Yang, C.-H. & Shapiro, R. 1973. The effects of the observational system and the method of interpolation on the computation of spectra. *J. Atmos. Sci. 30*, 530–536.

Yang, C.-H. 1974. A controlled experiment with one-dimensional interpolation. *J. Appl. Meteor. 13*, 625–636.

ТОЧНОСТЬ СПЕКТРАЛЬНОГО ПРЕДСТАВЛЕНИЯ ДАННЫХ С ПРОПУСКАМИ

Выбрана одномерная область, определяющая период аргумента четырех физических переменных, известных точно в 72 равномерно распределенных точках. Проведено 19 экспериментов с различным числом точек наблюдения, что соответствует различным пропускам в наблюдениях. Используя тригонометрические функции, как базис, построены функции Грама-Шмидта со свойствами ортогональности, отвечающими совокупности неравномерно распределенных точек. Для каждого допустимого масштаба свойства коэффициентов Фурье оцениваются для различных экспериментов в терминах обычных коэффициентов Фурье. Результаты показывают, что точность масштабных коэффициентов быстро ухудшается для масштабов, меньших, чем удвоенная величина максимального пропуска. Кроме того, при добавлении наблюдений в интервале, где имеются пропуски, наибольшая точность достигается путем минимизации наибольшего результирующего пробела.