#No 2017

(I.A. 4956/1970)

2017

No. 2159

Z. O. CUMALI
Member ASHRAE

Spectral Analysis of Coincident Weather Data for Application in Building Heating, Cooling Load and Energy Consumption Calculations

INTRODUCTION

Thermal loads on a building environmental system are greatly influenced by a number of weather variables. Dry-bulb temperature (dbt), net solar radiation intensity, dew point temperature (dpt), atmospheric pressure and wind velocity are some of these variables. To obtain a realistic estimate of the loads on a building the coincident patterns of dbt, dpt and solar radiation especially need to be considered. The set of algorithms compiled recently by ASHRAE and proposed for computer calculation of thermal loads and energy consumption estimates requires the use of coincident hourly weather data. It appears that the commonly used forms of weather data such as outdoor design temperatures, degree days, temperature level tabulations in cumulative hrs/season will not be adequate in fulfilling the newly created need for coincident data over extended periods of time.

One alternative is to use hourly weather data available from Environmental Science Services Administration (ESSA) recorded on magnetic tape, cards or in printed form. The problems associated with these records are the extensive editing and decoding necessary to convert them into usable form. The printed form, in addition, requires to be put into a computer readable format. Perhaps the more important problem is the uncertainty inherent in selecting a representative period from the available data records, which extend over many years. A rigorous statistical analysis on weather variables that results in a mathematical representation of their coincident behaviour would provide an acceptable solution to these problems. The study reported herein is concerned with the feasibility of such an approach and its effects on thermal load computation methods.

A number of studies have been made on weather data to determine statistical parameters of interest for thermal environmental engineering applications. The studies of Thom[1] and Holladay[2] are essentially concerned with the determination of design dbt. They have obtained a range of extremal values from many years of hourly temperature records and determined the probability distribution of the extremes. Their approach is essentially the same as that used in determining the published ASHRAE design temperature data. Crow[3] has extended this method to cover areas where only once per day recordings of weather data are made.

Hourly records of weather variables exhibit distinct patterns, long term annual patterns and short term diurnal patterns and random nondeterministic components. Bingham's[4] work indicates that the annual trend of temperature, based on weekly average of the daily mean temperatures and its time dependent variance can be adequately represented by 3 term Fourier Series. Jones[5,6] suggests methods for estimating the random component of daily aver-

Z. O. Cumali is Director, Consultants Computation Bureau, San Francisco, Calif. This paper was prepared for presentation at the ASHRAE Annual Meeting Kansas City, Missouri, June 28 – July 1, 1970.

ages using spectral methods and extends this to the multi-variate case. Kusuda and Achenbach[7] have found that the daily trends of dbt and dpt are expressible quite accurately at most a 4 term Fourier Series.

It is seen from the above that a detailed study of the correlation of weather variables which is of interest in thermal environmental engineering application has not yet been undertaken.

SPECTRAL ANALYSIS AND DIGITAL FILTERING

The inherent periodicity of weather data, as exemplified by the annual and diurnal cycles, is best observed by an analysis of its frequency composition. The transformation of the time records into a frequency representation is obtained by the use of the Fourier Transformation.

Considering any set of observations sequenced in time series, such as hourly recordings of dbt, let x(t) and y(t) be 2 time series defined over discrete values of t and of length N

$$\{x(t) \mid 0 \leq t \leq N-1\}, \{y(t) \mid 0 \leq t \leq N-1\}$$

The Digital Fourier Transform (DFT) of x(t) is defined as

$$X(n) = \sum^{N-1} x(t) e^{-i2\pi nt/N} \qquad (1)$$

and its inverse is

$$x(t) = \frac{1}{N} \sum_{N=0}^{N-1} X(n) e^{i2\pi nt/N} \qquad (2)$$

which is the discrete analog of the well known Fourier Transformation.

The transformed series terms, denoted by capital letters, X(n) are complex quantities and except for the factor N are equivalent to the complex form of the Fourier Series coefficients. The indices n denote the frequency in units of $2\pi/N$ radians.

An estimator for the power spectrum $S_x(n)$ of the variable x(t) is

$$S_x(n) = X(n)\, X^*(n) \qquad (3)$$

where * indicates the complex conjugate and is referred to as the periodogram estimate.

If we think of a time series, say dbt, as the sum of sine and cosine waves of distinct amplitudes and frequencies, the spectrum estimate is a measure of the power (the square of the amplitude) of each frequency component.

Measures of correlation between 2 time series are the cross-spectrum, or the coherency spectrum, and the phase spectrum. The cross spectrum is similarly defined as

$$S_{xy}(n) = X(n)\, Y^*(n) \qquad (4)$$

and is a complex valued quantity.

The coherency spectrum is based on the fact that the maximum attainable power at a common frequency is given by the geometric mean of the spectra of the 2 time series. The ratio of the cross spectrum to the geometric mean indicates the power actually attained relative to the maximum possible. Thus a definition of coherence is

$$C_{xy}(n) = \frac{|S_{xy}(n)|^2}{S_x(n)\; S_y(n)} \qquad (5)$$

which ranges in value between 0 and 1. If coherence is unity, the series are fully correlated in amplitude, if it has 0 value, they are uncorrelated. The phase spectrum indicates the phase lag or lead of the frequency components of the series and is given by

$$\theta_{xy}(n) = \tan^{-1}\left[\frac{-\mathrm{Im}(S_{xy}(a))}{\mathrm{Re}(S_{xy}(n))}\right] \qquad (6)$$

A fixed time lag between 2 periodic series would give

$$\theta_{xy}(n) = 2\pi n T \qquad (7)$$

where T is the time lag. This would be a straight line on the phase spectrum graph.

To observe the various frequency ranges in more detail, e.g., diurnal range, the variable records may be filtered, which amounts to their weighting with a prescribed set of constants. The filtering operation is a convolution operation

$$\bar{x}(t) = \sum_{K=-m}^{t=m} a(m)x(t-m) \qquad (8)$$

where $\bar{x}(t)$ is the filtered variable and a(m) are the 2m+1 filter constants.

This corresponds to filtering in the time domain. Filtering in the frequency domain is a simpler operation as the dft of Eq (8) gives

$$\bar{X}(n) = A(n)\ X(n) \qquad (9)$$

which indicates that convolution in the time domain corresponds to term by term multiplication in the frequency domain.

A very efficient algorithm was used in implementing the dft due to Sande[8] known as the Fast Fourier Transformation (FFT) both for spectral analysis and filtering of the data.

DATA ANALYSIS – PROCEDURES

Wheather data was obtained from ESSA tapes, series 140, for wind velocity, dbt, dpt, atmospheric pressure and total cloud cover. Solar radiation data was taken from ESSA, series 280, tapes. The period of record on one tape covers a 10-yr span of hourly values.

The records in all cases were found to have missing values and some magnitude errors in the variables. The magnitude errors were corrected by putting absolute value and maximum variation checks and substituting the average of the 2 adjacent data points. The missing data points did not exceed 12 consecutive hours in any variable. When more than 2 consecutive points were missing a substitution was made by interpolation of the adjacent 24 hr sequences to fit the missing part of the daily cycle after adjustment for the variable levels.

In order to look at the diurnal, seasonal and annual variations of the variables we have filtered each variable into 6 frequency ranges. The filter used for this purpose was one proposed by Jenkins.[9] Fig. 1 shows the gain of this filter for the various lengths used.

The filter is a symmetric high pass filter of length 2m+1 terms and we refer to m as the filter length. The filter coefficients are defined

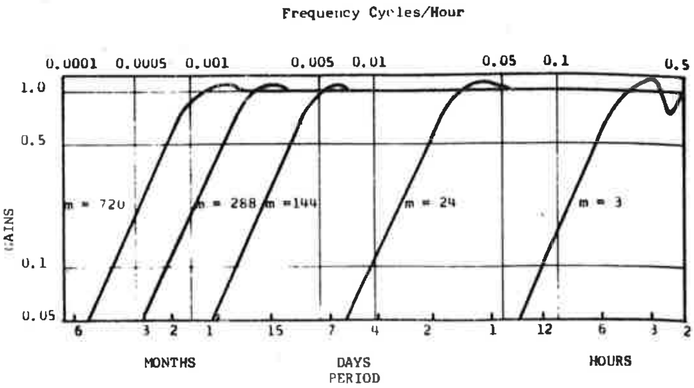

Fig. 1 Filter Characteristics for Various Lengths

$$\lambda_o = 1 - \frac{1}{m+1}$$

$$\lambda_j = -\left(1 + \cos\frac{\pi j}{m+1}\right) / (2m+2) \qquad j > o \quad (10)$$

$$\lambda_j = -\lambda_{-j}$$

and because of its symmetry, it does not result in any phase change. Its gain R for a given frequency n is the dft of $\{\lambda_j | o \le | j | \le m\}$

$$R(n) = \lambda_o + 2 \sum_{j=1}^{m} \lambda_j \cos 2\pi nj \qquad (11)$$

The filtering was applied sequentially to obtain the various frequency ranges.

If x(t) is one of the variables and $x_i(t)$ is the resulting series after applying a filter of length m_i for i=1 .... r, the series representing various frequency ranges are obtained by the following procedure:

$$\bar{x}_o(t) = x(t) - x_1(t)$$
$$\bar{x}_1(t) = x_1(t) - x_2(t)$$
$$\cdots\cdots\cdots\cdots$$
$$\bar{x}_i(t) = x_i(t) - x_{i+1}(t) \qquad (12)$$
$$\cdots\cdots\cdots\cdots$$
$$\bar{x}_r(t) = x_r(t)$$
$$m_i \cdot m_{i+1}$$

where $\bar{x}_i(t)$ contains the frequency range between the 2 filters of length m_i and m_{i+1}.

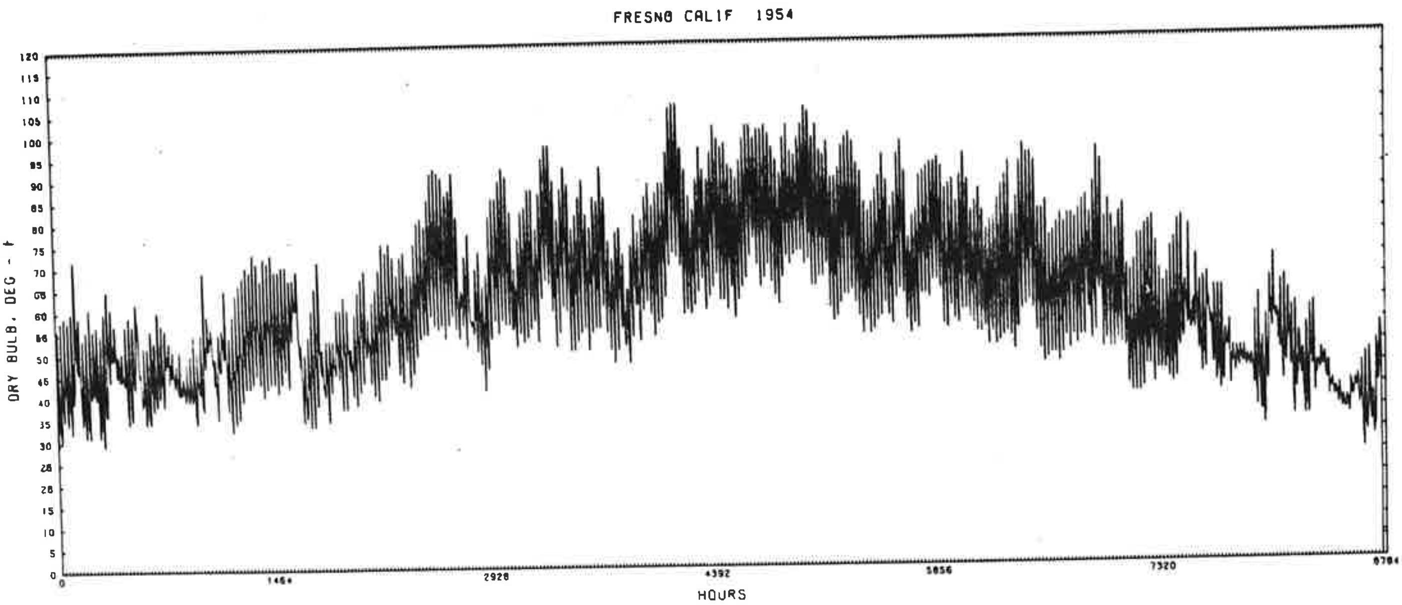

Fig. 2 Hourly Data — 1954.

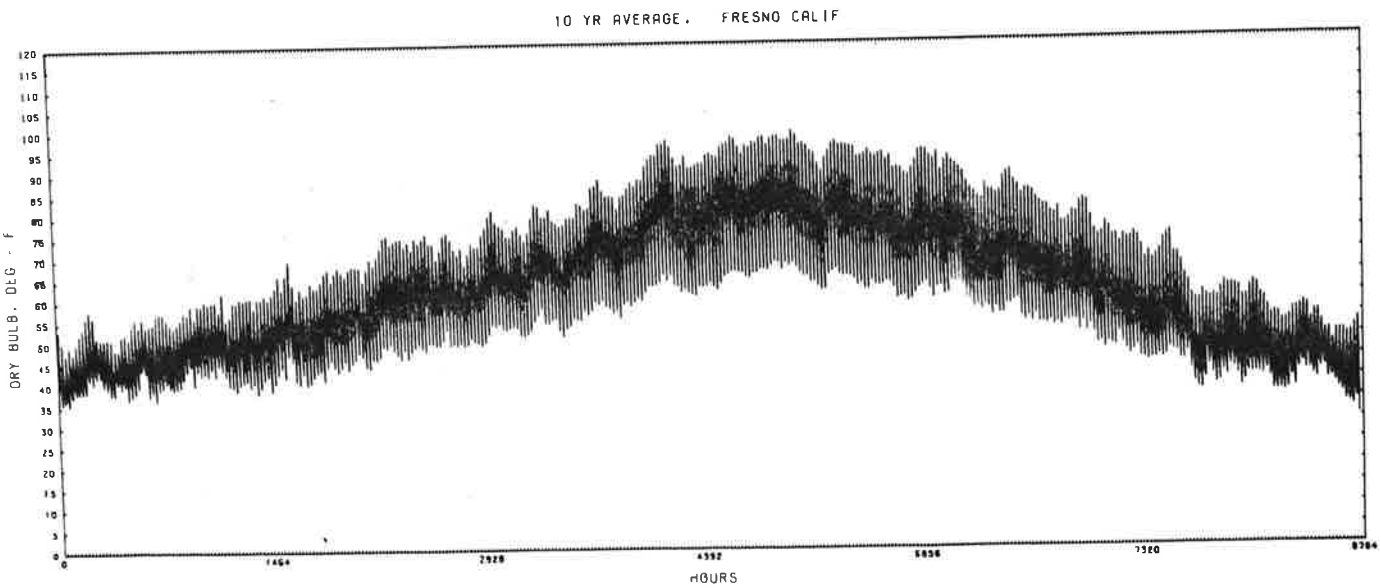

Fig. 3 Ten-Year Average Hourly Data.

Each variable was filtered in lengths of 720 hourly values corresponding to a 30-day span which is referred to as the monthly data. Two filters of length 3 and 24 were applied resulting in 3 frequency ranges. The low range of the monthly data has all the diurnal and higher frequencies filtered out of it. The diurnal range corresponds to the diurnal frequency and its first 4 harmonics, i.e., to 120 cycles/month. The high frequency range contains the residual high frequencies up to the Nyquist frequency of 360 cycles/month. Figs. 6 and 7 were obtained as a result of this sequential filtering. The upper frame is the plot of the actual hourly data with the low frequency range superimposed on it. The middle frame is a plot of the diurnal frequency range and the lower frame shows the residual high frequency range.

Two stages of filtering were applied to yearly and 10-yearly data. Each variable, in this case wind velocity, dbt and dpt and atmospheric pressure, was

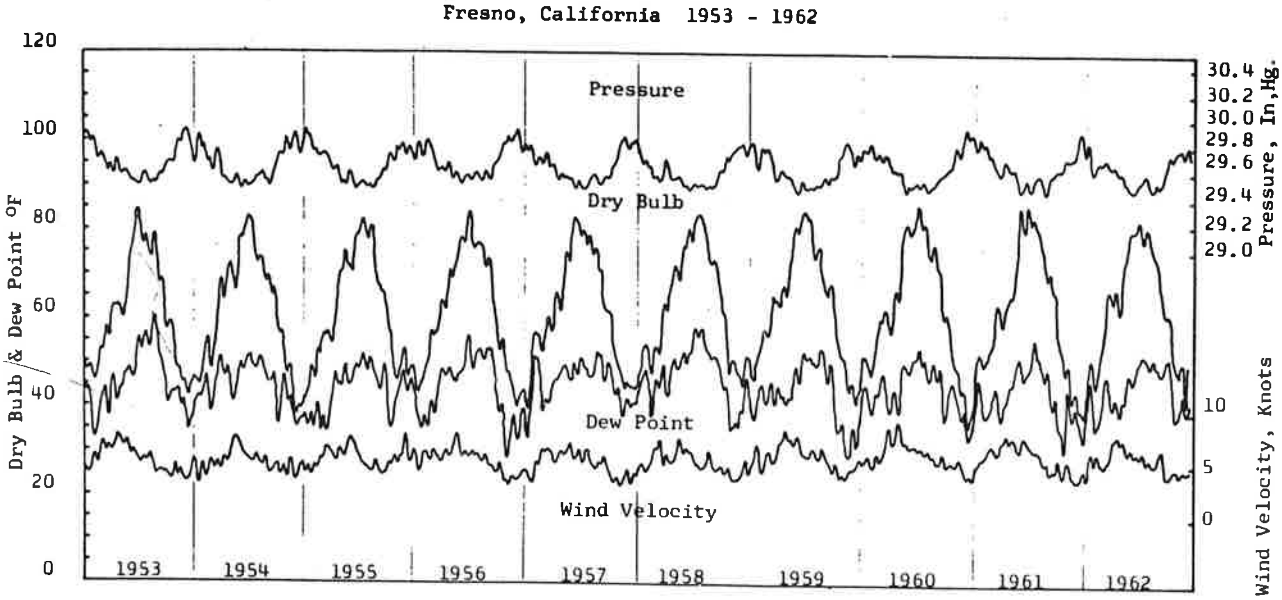

Fig. 4 Yearly Variable Patterns

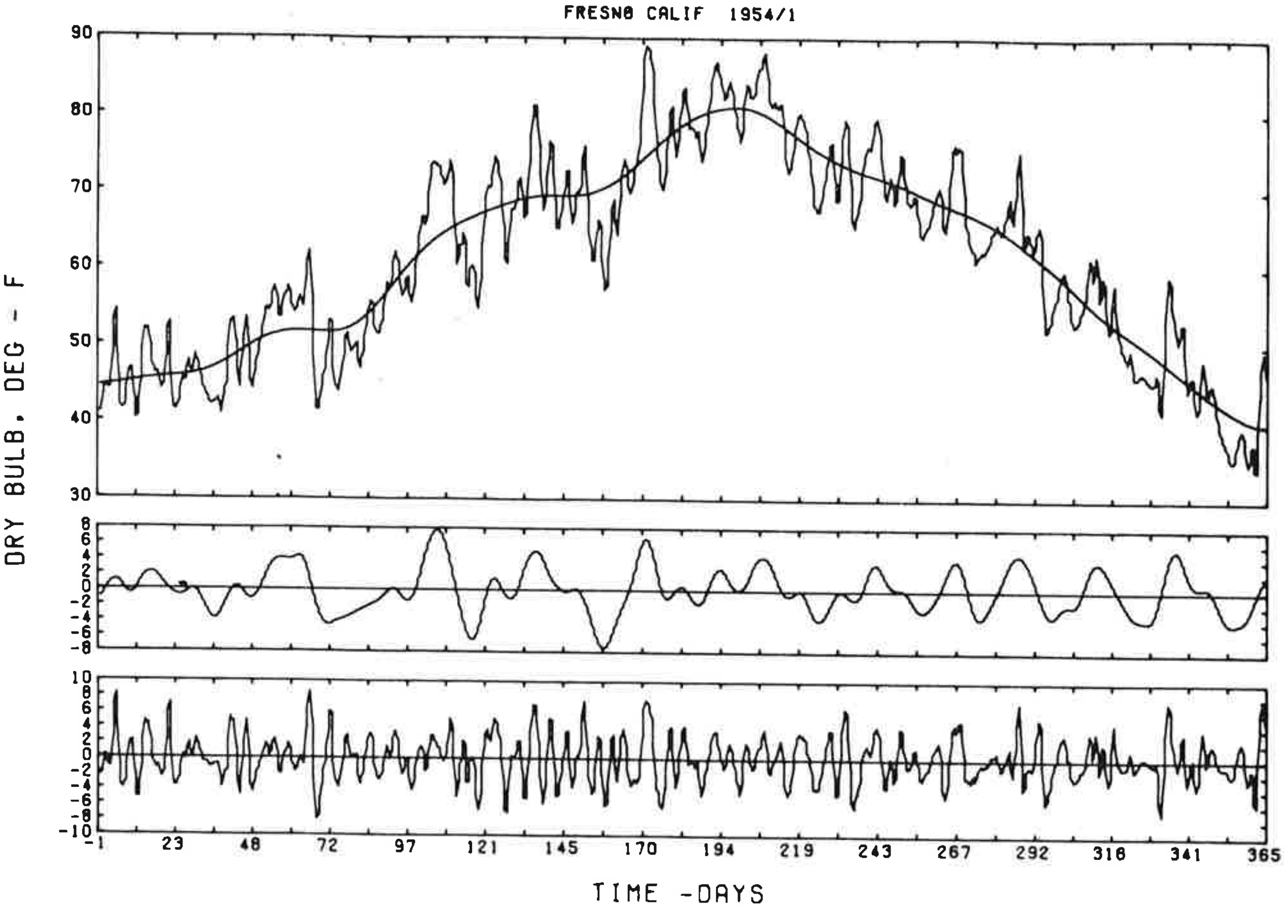

Fig. 5 Low Frequency Yearly Data – 1954.

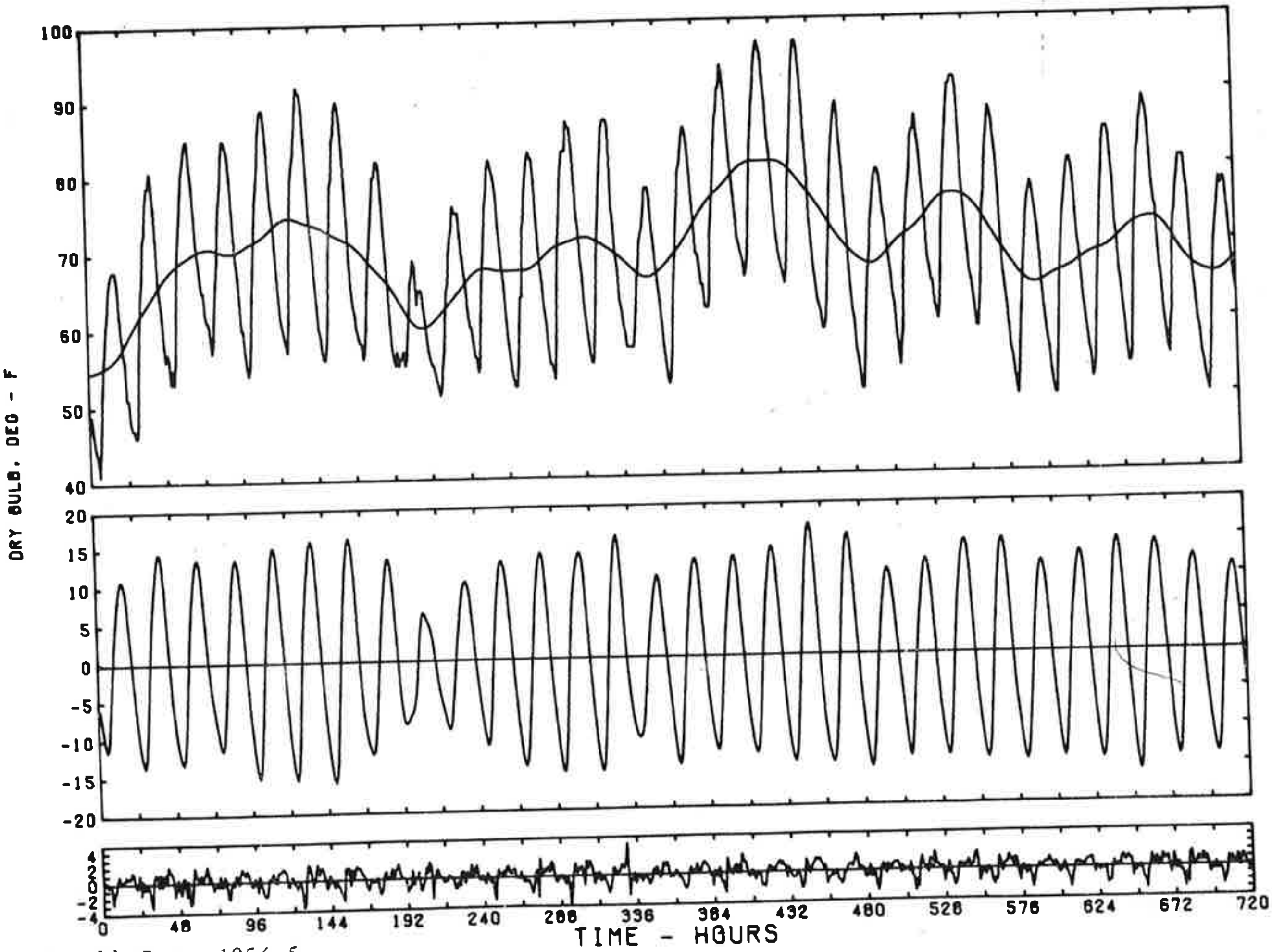

Fig. 6 Monthly Data — 1954–5.

filtered with a filter of length 24 to remove the diurnal and higher frequencies for 87,600 hourly values. The resulting filtered record was sampled at every 12th hr to produce a new record of 7,300 points. Each year in this record corresponds to 730 points. The resampling was done to reduce the calculation effort as the high frequencies are already taken out and this does not result in any loss of accuracy or detail in the data. The resampled data is referred to as the low frequency data in what follows. The low frequency was further filtered with a filter of length 24, corresponding to a filter length of 288 on hourly data, the remaining lower frequency range which contains only 2 cycles or less/month is shown in Fig. 4, all 4 variables superimposed on one frame. Here the yearly cycles are clearly identifiable.

Another sequence of filters was applied to 1 year, 730 points, of the 10-year low frequency data. Filters of length 12 and 60 were used. Fig. 5 shows plots of the resulting frequency ranges. Fig. 5 has in the upper frame the low frequency data for 1 year plotted on an enlarged scale. Superimposed on it we see the frequency range up to 6 cycles/year corresponding to the yearly cycle and its few harmonics. The center frame shows the monthly frequency range, 6 to 24 cycles/year.

The lower frame is the residual containing the 24 to 360 cycles/year. This we call the weekly frequency range as the main prominences have 4 to 10-day periods.

In order to compare the yearly variation of the hourly data records 10-yr averages were computed for 4 variables. Fig. 3 shows the 10-yr average hourly data for dbt. It may be compared to the hourly data for the year 1954 shown in Fig. 2.

In computing the estimates of spectra for the variables in various frequency ranges we have followed mainly the methods suggested by Bingham, Godfrey and Tukey.[10] In all cases we used data lengths of 720 points, which have the factors $5 \times 3^2 \times 2^4$ that are symmetric about the factor 5.

This was chosen for computational convenience in using Sande version of the FFT even though this length corresponds to 12⅙ 30-day months in a yr, or a 360 day year for the low frequency data.

The data was first reduced to 0 mean by computing the sample time average and subtracting it from the 720 terms. Then the FFT program was used to compute the dft. In order to eliminate the end effects due to the sectioning, and leakage, the dft coefficients were "hanned", i.e., convolved with -¼, ½, -¼ weights.

The spectral estimate was obtained using Eq 3, and the number of terms computed varied in accordance with the particular frequency range of the data up to n = 360.

The smoothing of the initial spectrum estimate was done by using a triangular filter with ¼, ½, ¼ weights. Graphs of the variable spectrum estimates shown in the 2 upper frames of Figs. 9 through 12 have been smoothed 4 times. All spectrum estimates are plotted on logarithmic scale with a linear frequency axis and only the top 4 decades are shown.

The estimates for the coherence and phase spectra shown in the lower 2 frames of Figs. 9 through 12 were obtained by computing the dft of 2 variables x(t) and y(t) and hanning them. Then the cross spectrum estimates were computed using Eq (4) and smoothed 4 times with the triangular filter. The coherence and phase spectra estimates were obtained from Eqs (5) and (6).

RESULTS

The weather data analysed was limited to one station, Station #93193 for Fresno, California, for the years 1953–1962. The variables considered in detail were wind velocity (VW), dry bulb temperature (dbt), dew point temperature (dpt) and atmospheric pressure (P). Total solar radiation intensity (S) and total cloud cover (C) were also considered but not in as much detail.

One of the interesting characteristics of the recorded solar data for this station was that values obtained from the standard formula used for solar radiation intensity in building load calculations given in the 1967 ASHRAE GUIDE and DATA BOOK did not compare very favorably with it. Especially in the mornings, the values computed by the formula were quite low, and in the afternoons the opposit occurred. Also the maximum values differed approximately 20%. The morning and afternoon differences ranged from 49 to 20% for clear days. Originally, it was our intention to use the solar radiation formula as a basic predictor, i.e., the only variable which was available to formulation in a direct way. Unfortunately, with this much variability discovered, it was decided to use solar data on the records as is and leave the search for the reasons of this variation for a future study. This was the basic reason for studying the other 4 variables in more detail. The emphasis on cloud cover data was also relaxed as its relation to solar data is the most important one.

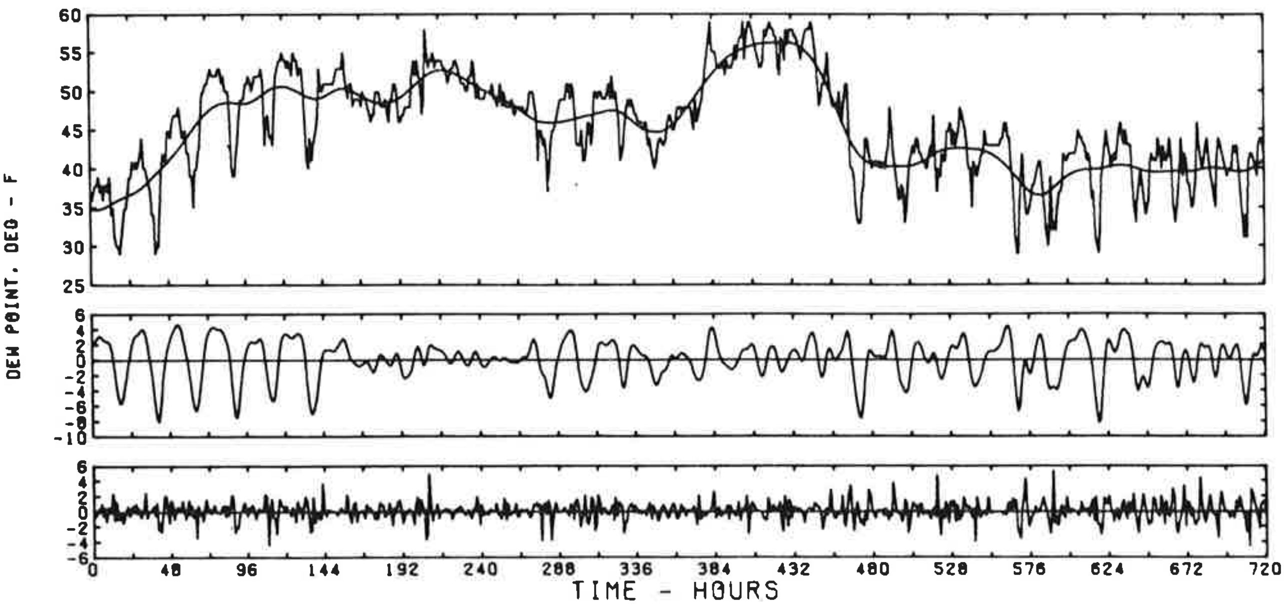

Fig. 7 Monthly Data – 1954–5.

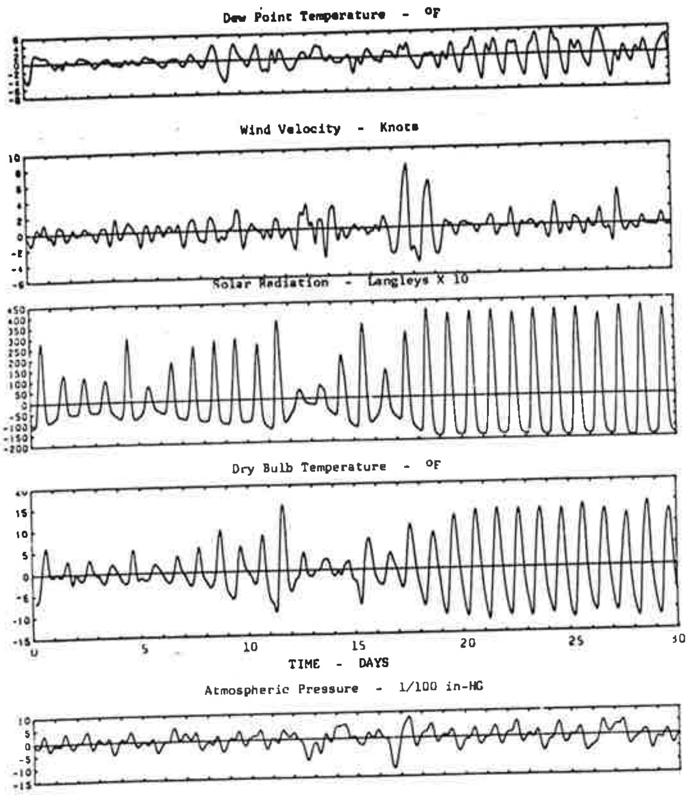

Fig. 8 Diurnal Cycles of Five Variables

Approximately 400 graphs were generated in the course of the study. The ones included in the paper are representative of the observed trends. Fig. 2 shows the hourly dbt for the year 1954. The yearly trend of temperature and the superimposed variations from the average trends are apparent in comparison to Fig. 3 which shows the 10-yr average of the hourly dbt. Fig. 4 shows the 0 – 24 cycles/yr range of the annual trends of 4 variables. dbt, dpt and WV have distinctly correlated trends and the P is inversely correlated with them. The coherences in the 1 – 6 cycles/yr were in all cases above 0.90 for the first harmonic and decreased somewhat linearly to 0.50 at the 6th harmonic. It should be noted that the coherency as reflected by Eq (5) is proportional to the square of the amplitudes. Amplitude correlations are thus proportional to the square root of the coherence. It is assumed that coherences less than 0.5 are considered as an indication of no correlation or randomness from a practical point of view.

Fig. 5 shows a further frequency decomposition of the dbt in the range 0 – 360 cycles/yr. The sum of the center frame and the smooth curve in the top frame corresponds to the 1954 dbt curve in Fig. 4.

The coherency and phase spectra of the 6 – 24 and 24 – 360 cycles/year ranges for dbt and dpt are shown in Figs. 9 and 10. It is interesting to note that in the 6 – 24 cycle range the peak coherence of 0.85 is at 12 cycles/year. This was observed to be true in all the variables except cloud cover. The range 24 – 360 cycles. Fig. 10, has 2 peaks at 30 and 50 cycles/year with coherences of 0.88, the latter corresponding to an approximate weekly period. This range although showing high amplitude correlation among the variables and consistent phase lead or lag trends is smoothed out on averaging, e.g., Fig. 3. As this range determines in effect the year to year variation of the seasonal variable trends, it is not deterministic and predictable only in a statistical sense. However, once a pattern is prescribed on a variable, the other variables obey the casual relationships among them resulting in the observed correlations. Jones'[5] findings on the predictability of daily average temperatures supports this point of view.

In the diurnal range 30 – 120 cycles/month, the correlations were found to be quite consistent and high as was expected. Figs. 6 and 7 show the frequency decomposition of this range for dbt and dpt. The upper frames show the plot of the recorded hourly values and the superimposed curve is the filtered low frequency trend. The diurnal range, the daily cycle and its first 4 harmonics, or 30 – 120 cycles/month is shown in the center frames. The spectra in this range, upper frames in Figs. 11 and 12, show the importance of the first harmonic, which is 5 to 20 times the magnitude of the second harmonic excepting dpt and cloud cover. This indicates that 2 to 3 harmonics would describe quite adequately the daily patterns of the variables. The coherences are throughout in the range 0.9 to 1.0 for the first 2 harmonics which is indicative of the very substantial correlations among the variables. The phase spectra show the commonly observed lead and lag relationships, i.e., dbt lagging S and dpt lagging S and dbt etc. The well defined correlations in the diurnal range of frequencies are very promising in defining simultaneously the patterns of a number of variables for use in design load calculations. In Fig. 8 filtered diurnal course of the variables are shown for the month of May, 1954.

The high frequency residual 120 – 360 cycles/month shown in the bottom frames of Figs. 6 and 7 appears to be random as we have not observed any

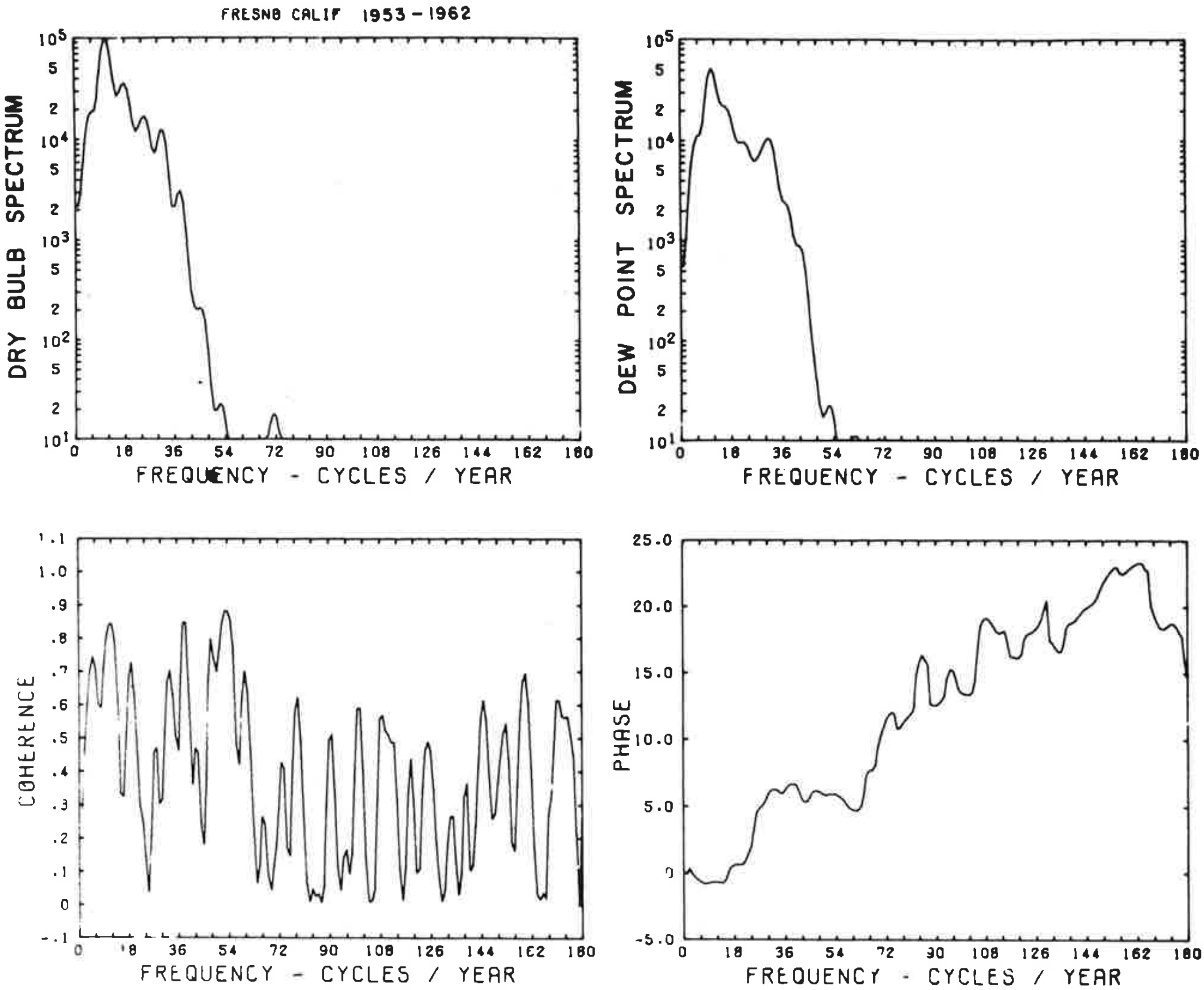

Fig. 9 Yearly Data Cross Spectra — 2.

consistent correlation patterns or predominant frequencies. Its magnitude is small relative to the diurnal cycle amplitudes.

Two interesting phenomena were also observed. Fig. 13 shows the envelope of the diurnal amplitudes in dotted lines, which suggests a form of amplitude modulation. Comparison with the smooth trend curve in the upper frame shows an inverse correlation. This type of correlation was observed visually for most of the variables but was not analyzed in detail. A technique known as complex demodulation will be used to examine this phenomenon in the next phase of the study. Another observation made was the seasonal changes in the variance of the ranges 24 – 360 cycles/year and the 120 – 360 cycles/month. Especially for atmospheric pressure variable the variances were found to be high in the winter and low in the summer seasons. Bingham's[4] work shows that this is the case for the dry bulb in the range of frequencies less than the diurnal.

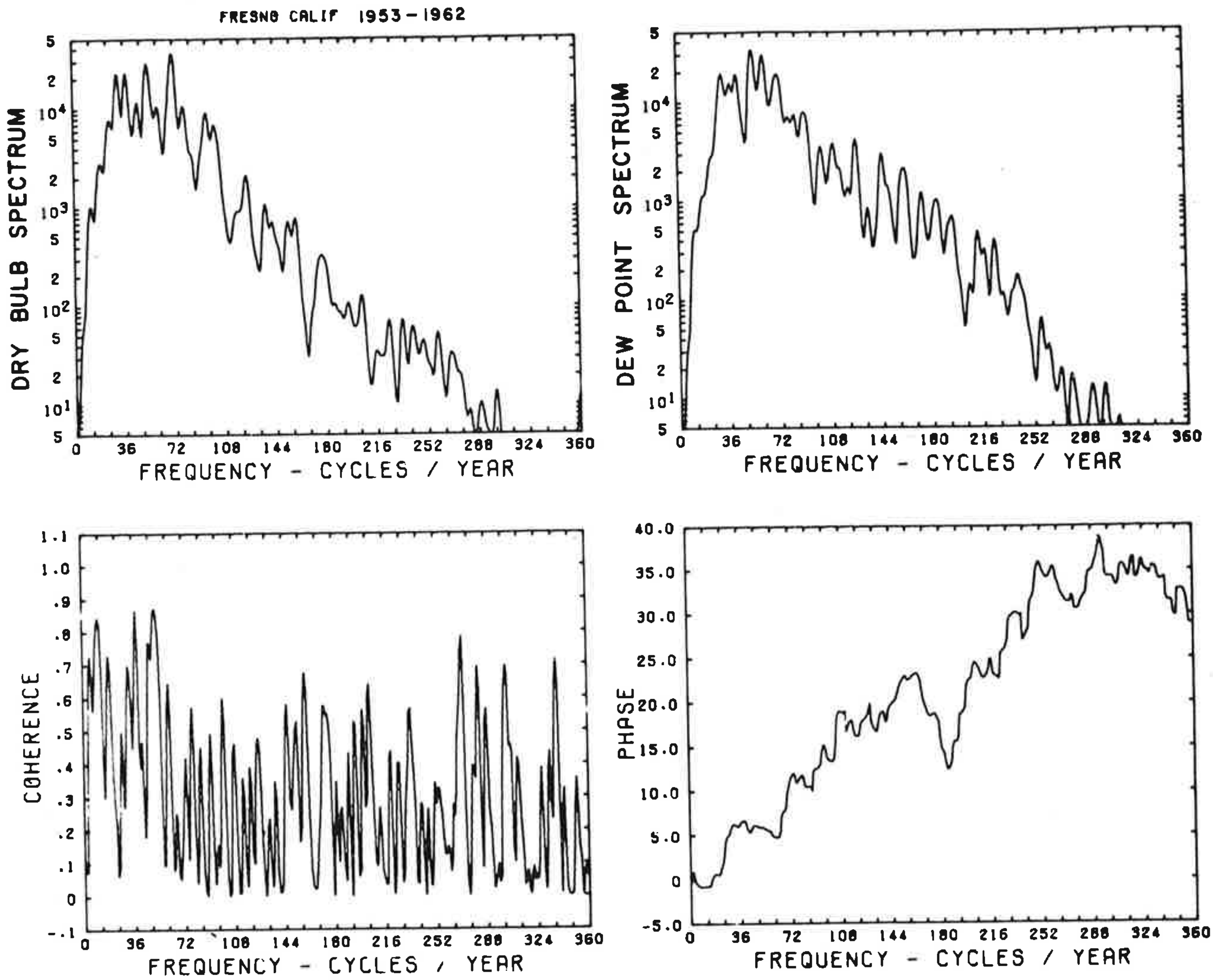

Fig. 10 Yearly Data Cross Spectra – 3.

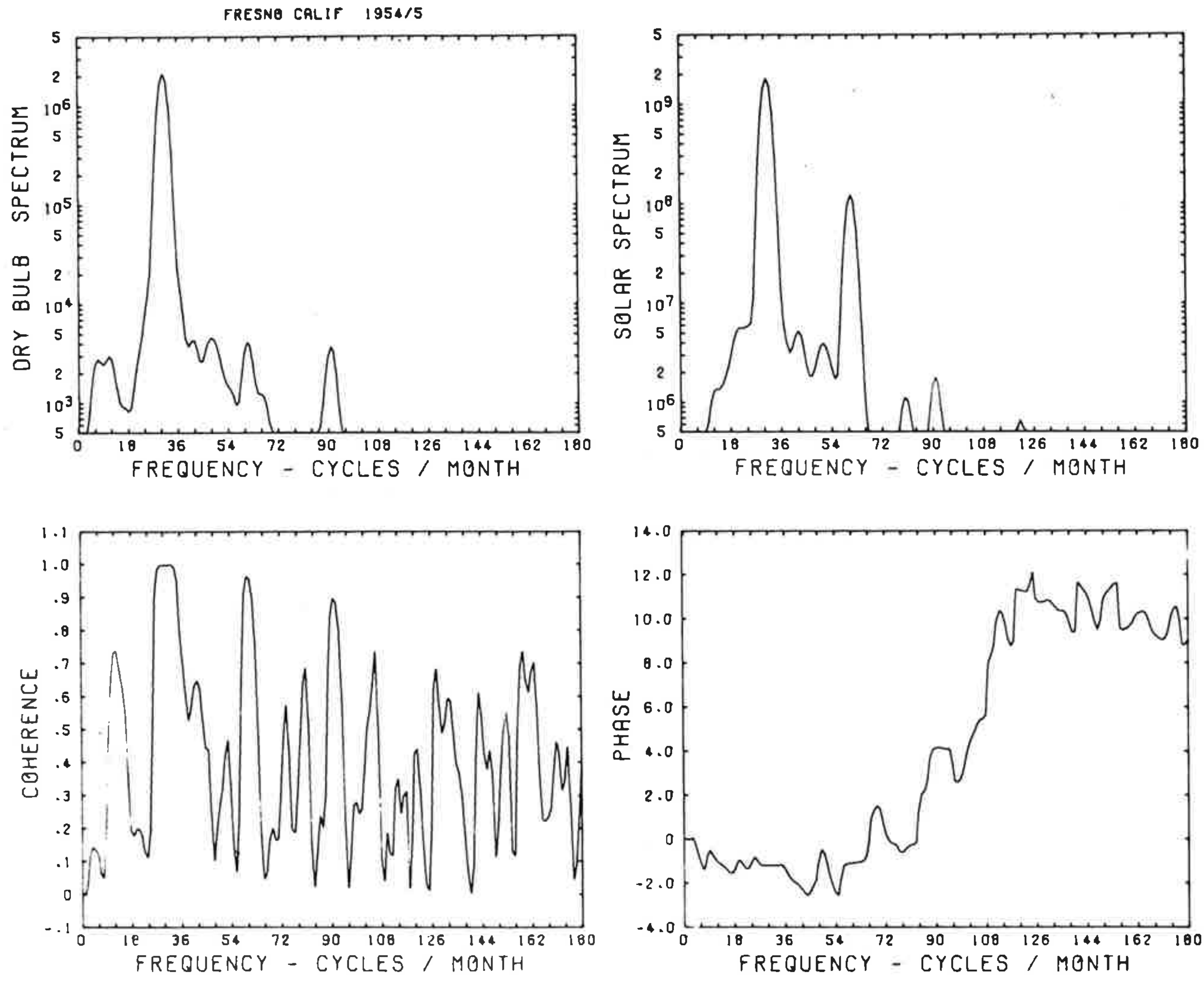

Fig. 11 Diurnal Frequency Range Cross Spectra.

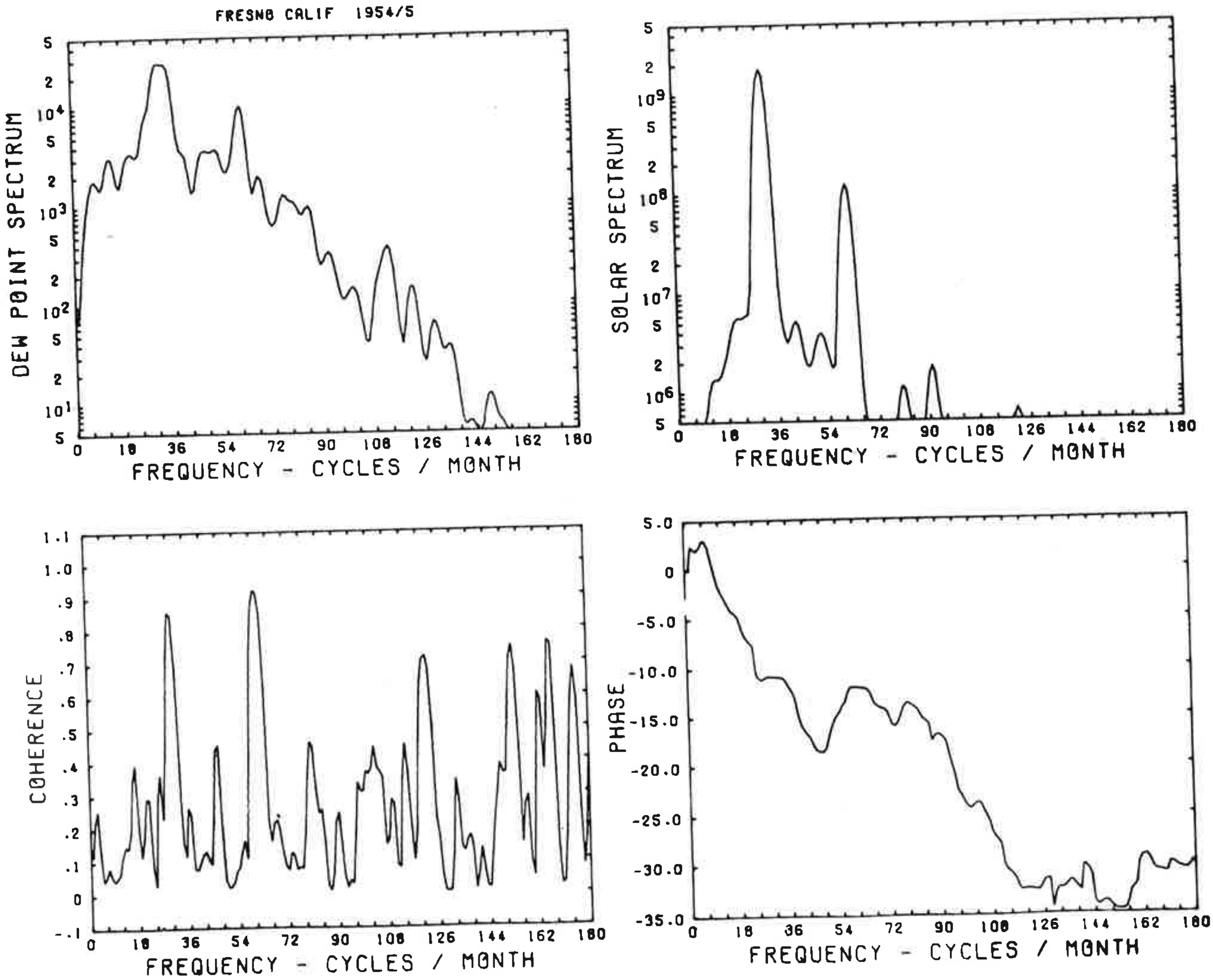

Fig. 12 Diurnal Frequency Range Cross Spectra.

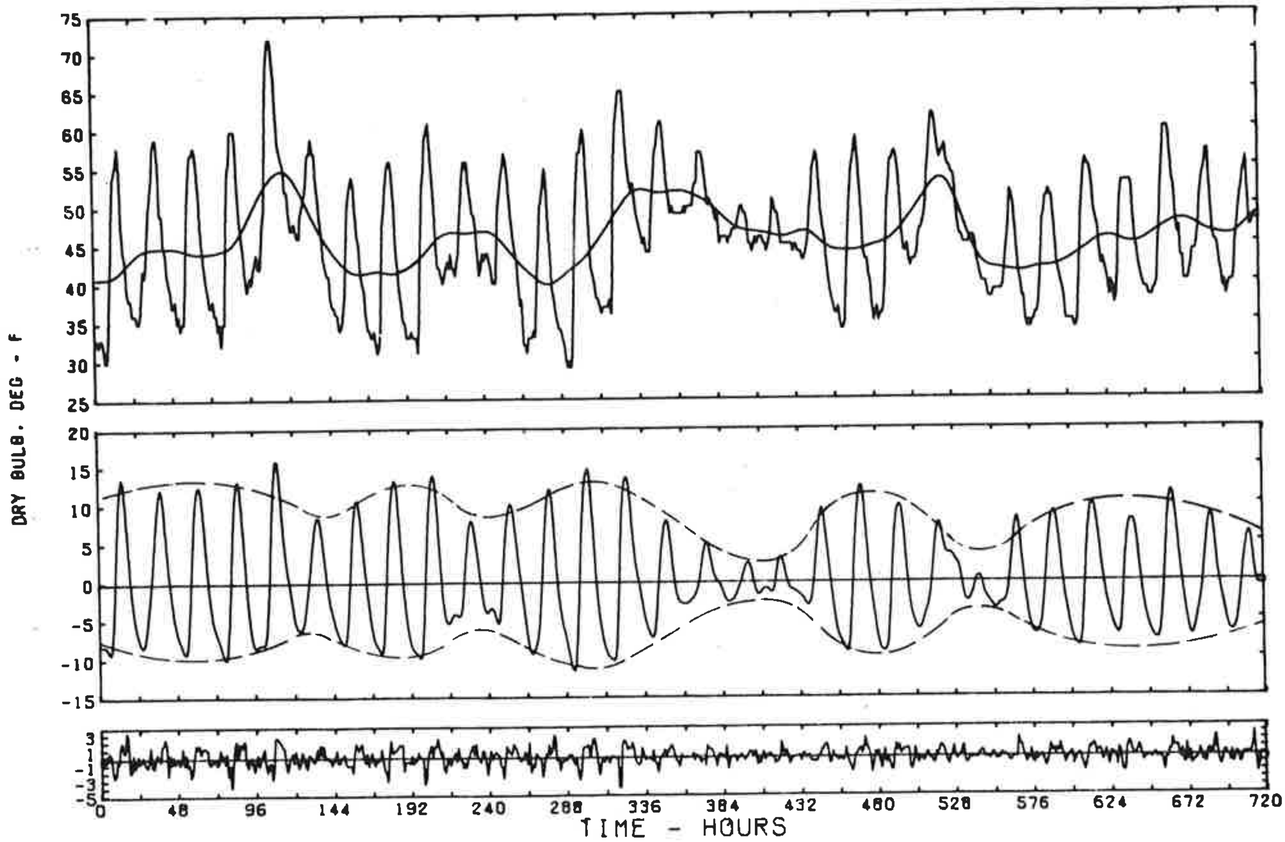

Fig. 13 Monthly Data—1954—2.

APPLICATION AND FUTURE WORK

A simple application of spectral analysis of weather data is the estimation of average coincident patterns. If x(t) and y(t) are dbt and S respectively for example, an estimate of the average coincident diurnal pattern of y(t) can be obtained if the x(t) is given. Assuming that the causal relationship between x(t) and y(t) is linear (i.e., a linear integral or differential equation) their spectra and cross spectrum are related

$$S_{xy}(n) = S_x(n)\,R(n) \tag{13}$$

$$S_y(n) = S_x(n)\,R(n)\,R^*(n) \tag{14}$$

where R(n) is the complex gain of this system. Therefore, from the cross spectrum $S_{xy}(n)$ and the spectrum of x(t), $S_x(n)$ we obtain

$$R(n) = S_{xy}(n) \;/\; S_x(n) = |\,R(n)\,|\; e^{i\phi(n)} \tag{15}$$

If x(t) is expanded in a Fourier Series

$$x(t) = a_o + \sum_{n=1}^{m} a_n \cos\left[2\pi nt/N + \theta(n)\right] \tag{16}$$

the terms in the Fourier Series expansion of y(t)

$$y(t) = b_o \quad \sum_{n=1}^{m} b_n \cos\left[2\pi nt/N + \Psi(n)\right] \tag{17}$$

are obtained from the a(n) and the R(n)

$$b_n = a_n\,|\,R(n)\,| \tag{18}$$

$$\Psi(n) = \theta(n) + \phi(n)$$

The length of the Fourier Series m is at most 3, as Fig. 10 spectra indicates 1 large and 2 smaller distinct peaks. Also note that ratio of the peak values of the spectra at frequency n is $|R(n)|^2$. The phase angle $\Phi(n)$ is the slope of the phase spectrum of frequency n. The other coefficients a_o, b_o, are the average values of x(t) and y(t) over the period of interest, i.e., 1 month.

This example may be extended to the multivariable case to estimate the coincident patterns of all 6 variables. Details of this approach are given by Jenkins[9] and Robinson.[11] As some weather variables are better predictors than others, e.g., S and P, it is apparent that multivariate estimate of coincident patterns are an improvement over bivariate ones.

As weather data is not stationary, e.g., averages and variances change with the seasons, spectral analysis is applicable over limited periods of time, say 1 month or less, for coincident diurnal pattern estimates to assure that the segmented record is approximately stationary. Jones[6] has suggested a method which overcomes this limitation. As weather data is periodic in the mean square sense, it may be expressed in an auto-regressive form with periodic coefficients.

$$x(t) = \sum_{m=1}^{n} a(t-m)\, x(t-m) + e(t) \qquad (19)$$

where $\hat{x}(t)$ is the estimated value in terms of the past values of the variable x(t) and e(t) is the error term. The coefficients a(t-m) are periodic with periods corresponding to the diurnal and annual cycles and their harmonics. Using the multivariate form of Eq (19) the coincident diurnal patterns of the variables can be estimated at any time of the year. The application of this method to weather data will be reported in a future article.

The problem of selecting a representative year, with the need for coincident hourly records of 6 or more variables for load calculations and system energy consumption estimates, is also dependent somewhat on factors other than weather data. A review of the proposed load calculation methods illustrates this point.

If the total load on a zone is given by q(t) it is the sum of loads due to each input, $q_i(t)$

$$q(t) = \sum_{i=1}^{m} q_i(t) \qquad (20)$$

where each $q_i(t)$ is given by

$$q_i(t) = \sum_{j=1}^{k} w_i(j)\, x_i(t-j) \qquad (21)$$

The input variables $x_i(t)$ are convolved with the weighting functions $w_i(j)$ corresponding to them. The annual energy requirement Φ is thus given by summing the heating and cooling loads separately

$$\Phi_c = \sum_{i=0}^{n-1} q^+(t) \qquad (22)$$

$$\Phi_n = \sum_{i=0}^{n-1} q^-(t) \qquad (23)$$

where the load functions q^+ and q^- are equal to 0, if they are negative and positive respectively otherwise they are equal to q(t). By taking the dft of $q_i(t)$ we obtain

$$q_i(t) = \sum_{j=1}^{k} w_i(j)\, x_j\,(t-j) \supset Q_i(n) = W_i(n)\, X_i(n) \qquad (24)$$

where $W_i(n)$ corresponds to the frequency response function of the particular zone with respect to the input variable x_i. The $Q_i(n)$ are equivalent to the FS coefficients except for the factor N, the number of hrs over which the loads are computed. Thus as w_i weight x_i over a number of past hrs to get the load $q_i(t)$, the frequency response function weights each frequency component individually.

The dft of q(t) is the sum of Qi(n) for the m input variables $x_i(t)$

$$Q(n) = \sum_{i=1}^{m} W_i(n)\, X_i(n) \qquad (25)$$

and the load function is simply the inverse DFT

$$q(t) = \frac{I}{N} \sum_{n=0}^{n-1} Q(n) e^{i2\pi nt/N} \qquad (26)$$

We note that the oth term of the DFT, Q(o), equals the net area under the load function, i.e.,

$$\sum_{t=0}^{n-1} q(t) = Q(o) \qquad (27)$$

This corresponds to the algebraic sum of Φ_c and Φ_n,

$$Q(o) = \Phi_c + \Phi_n \qquad (28)$$

in other words as $\Phi_c > o$ and $\Phi_n < o$, it corresponds to the difference in the magnitude of the annual cooling and heating energy requirements in Btus. In order to obtain Φ_c or Φ_n we need one additional value, namely

$$\Phi_T = \sum_{t=0}^{n-1} |q(t)| = \Phi_c - \Phi_h \tag{29}$$

Averaging the input weather data over the years at each hr corresponds to an ensemble average estimate, and summing and convolution operations are commutative with the averaging operation, i.e., expected value operator E. We may then write noting Eq (28)

$$\left.\begin{aligned} E[Q(o)] &= E[\sum_{n=0}^{n=1} q(t)] \\ &= E[\sum_{t=0}^{n=1} \sum_{i=1}^{m} \sum_{j=1}^{k} w_i(j)\, x_i(t-j)] \\ &= \sum_{t=0}^{n-1} \sum_{i=1}^{m} \sum_{j=1}^{k} w_i(j)\, E[x_i(t-j)] \end{aligned}\right\} \tag{30}$$

which shows that an estimate of the average value of the area under the load function q(t) may be obtained from the estimate of the ensemble average of the input variables. However, when we apply the above to

$$E[\Phi_T] = \sum_{t=0}^{n-1} E[|q(t)|] \doteq \sum_{t=0}^{n-1} E[| \sum_{i=1}^{m} q_i(t) |] \tag{31}$$

we note that

$$| \sum_{i=1}^{m} q_i(t) | < \sum_{i=1}^{m} |q_i(t)| \tag{32}$$

and the argument breaks down at this point. In other words, Φ_T is not a simple functional of the ensemble average estimates E[xi(t)] of the input variables. As Φ_c and Φ_n are functions of Φ_T, this applies to them as well.

It is apparent that data averaged hourly over the years cannot be used for energy consumption estimates. The summing of the individual loads qi(t), due to the input weather variables xi(t), which may be positive or negative independently of each other to some extent, results in this situation as shown in Equation (31). Therefore, the building characteristics do seem to have an influence in considerations of representative year selection. The inadequacy of the degree day approach, other than being a single variable approach to energy consumption estimates, can be attributed to the fact that it does not consider transient effects in contrast to the proposed procedures. It is obvious that more research is needed in this area.

Eqs (26–28) suggest that with the use of the dft significant saving in computational effort may be achieved in energy consumption estimates. For purely heating or cooling seasons net energy requirements may be obtained from Eq (28) as Φ_T is equal to Φ_c or Φ_h in this case.

CONCLUSION

Use of spectral analysis as a tool in identifying correlations of annual and diurnal patterns of weather data is presented. Results obtained with the application of this technique to the 10-yr hourly data for 6 variables of Fresno, Calif., in the 1952–1963 period indicate significant correlations in the annual and diurnal patterns of the variables.

A method is introduced for estimating coincident diurnal patterns, which can be used in load and energy studies. Considerations in selection of representative periods of weather data for standard usage are found to require further research.

REFERENCES

1. H. C. S. Thom, Revised Winter Outside Design Temperatures, TRANSACTIONS ASHAE 63, 111 (1957)
2. W. L. Holladay, Local Climatic Weather Data, TRANSACTIONS ASHAE 64, 229 (1958)
3. L. W. Crow, Study of Weather Design Conditions Research Project #23, LWC Report No. 40
4. C. Bingham, Distribution of Weekly Averages of Diurnal Temperature Means and Ranges about Harmonic Curves, Monthly Weather Review, 89, 357 (1961)
5. R. H. Jones, Spectral Analysis and Linear Prediction of Meteorological Time Series, J. Appl. Meteorology 3, 45–42, (1964)

6. R. H. Jones, Brelsford, W.M., Time Series with Periodic Structure, John Hopkins University, Dept. of Stat., Tech. Report 56, July, 1966
7. T. Kusuda, Achenbach, P. R., Coincident Weather Characteristics of Six Selected Cities in the United States, in Symp. on Weather, ASHRAE Annual Meeting, June 1966, Toronto, Canada
8. G. Sande, Gentleman, W. M. Fast Fourier Transforms for Fun and Profit. Proc. AFIPS, Fall Joint Computer Conference, 29, 563 (1966)
9. G. M. Jenkins, Watts, D. C., Spectral Analysis and its Applications, Holden Day, San Francisco, (1968)
10. C. Bingham, Godfrey, M. D., & Tukey, J. N., Modern Techniques of Power Spectrum Estimation, IEEE Trans. on Audio & Electro-Acoustics AU-15, 56 (1966)
11. E. A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Holden Day, San Francisco (1967)

ACKNOWLEDGEMENT

This work was done for the National Bureau of Standards Building Research Division, Environmental Engineering Section, and supported by the U. S. Post Office Department, Bureau of Research and Engineering, as a part of its Construction Research Program. The encouragement and suggestions of Dr. Tamami Kusuda of NBS throughout the study are gratefully acknowledged.

DISCUSSION

M. LOKMANHEKIM, (General American Research, Chicago, Ill.): Does the author believe that the relationship you find for Fresno, Calif., can be generalized for the continental United States?

MR. CUMALI: The methodology introduced here is applicable to any kind of information. Meaning it could be weather data from one station, it could be weather data from another station or it is also applicable to other data. For example, relationship of data from building to the load. But the methodology is general. We have done this only for one station. We are presently doing this for another station in Chicago using exactly the same methodology.

But as far as being able to pick up a certain number of stations across the country which are representative of the climatic conditions or regimes available in the United States, whether that can be done I really don't know. The only thing I can refer you to is that one gentleman at the University of Chicago has done an analysis to represent the weekly average temperatures. He has done a 4 term Fourier Series representation for the northeast U. S. and drawn curves of these coefficients on a map to show how they change. So it appears that for at least one variable this thing has been done and is applicable. Whether it is totally applicable remains to be seen, but at least there is an indication that it can be done for more than one variable.

MR. LOKMANHEKIM: In this manner, I am disagreeing with you because about 8 or 10 years ago I had done the very same thing, making a Fourier series analysis for Chicago for some special project. It seems to me that the mathematical procedures are O.K. but I believe you have to take into account the meteorological effect. You cannot make a judgment assuming that this method is true for every city. You have to use your meteorological judgment to get a good correlation. I have my doubt that this gives you a correlation, as you state, for every city in the United States.

MR. CUMALI: Oh, I did not say it would give a correlation to every city. I just said the method can be applied to the data. And as for using meteorological judgment, that is already –

MR. LOKMANHEKIM: No, what I am saying, for example, is that you find for Fresno a relationship between Fresno and solar radiation.

MR. CUMALI: Right.

MR. LOKMANHEKIM: But you are not going to find the same kind of relationship in Chicago or in Maine or in –

MR. CUMALI: I fully agree with you. The correlations change but the method of obtaining the correlations is the same. That is all I am saying.

MR. LOKMANHEKIM: How are you going to put the general mathematical relationship influenced by the local pattern? For example, we know that in the United States, Tucson and Phoenix, Ariz. are about 10 miles apart and they give two different climates. How are you going to use cities in your mathematical analysis?

MR. CUMALI: My mathematical analysis as you describe it to me is fairly general and it does not even claim that it can do that. It is not mine; I am only using it. It is the methodologist's task to do this kind of analysis. All I am trying to do is to make an analysis on a given piece of information. And I am sure there are certain basic correlations which can be used for many purposes. How they are put together is something else again.

MR. LOKMANHEKIM: Let me give you one example. About 6 years ago we had a Civil Defense award. We computerized a 10 year span of constant weather data and we built up the isometric lines for the continental United States. We plotted these full computerized. Then we took the data to a very prominent research scientist in the country. He looked and we saw him smile. He said, "You are wrong." He told us, "You are mechanical engineers, O.K. You did these things and it is perfectly all right, but I have to make some meteorological judgment and correct it." And he started playing around with cities and islands and so forth. I was wondering how he can do these things in this work. He said, "You can't represent this."

MR. CUMALI: The only way that something like this can be attempted is by doing this on a regional scale. For example, this is where you need the help of the meteorologist who is to identify what are the basic regimes in a given area, which areas would be representative of the overall climate in a grid type of thing. Now once you do this over a given area, perhaps you could try to define some relationships which can do this on an area basis. At this point, all we are doing is taking one station and making an analysis on that station. We have not looked at the other problem. We hope that if there are funds available and time available we would like to look at this in the future.

MR. LOKMANHEKIM: I have also another suggestion. As far as the weather data is concerned I, for example, eliminate the solar radiation from the picture. Instead of solar radiation, I go to some other climate because solar radiation data in the continental United States is very limited and is not generalized. I go to a cloud cover, a cloud type, and use a mathematical model for the calculation of solar radiation. That's my suggestion.

MR. CUMALI: I think that is a very valid suggestion because solar radiation usually comes on a separate tape. And as you say there are few stations in the United States who measure and record this.

MR. LOKMANHEKIM: And also you have to remember that solar radiation data taken in the stations and the weather record stations in the same location are not the same place. For example, in Chicago you have Midway, you have O'Hare and you have Glenview but the solar radiation data is taken at Adler Planetarium Laboratories, which are not close to each other. In this case, I suggest strongly that you eliminate the solar radiation, because I believe the method right now built up at the National Research Council and the method built up at NBS and the method built up at General American without taking the intensities of the solar radiation is perfect. If we are able to introduce this plot or modifier in the picture, we will be better off as far as getting correlation and it will be very useful I think.

MR. CUMALI: Thank you. I think that is a very valid comment.
