AN ANALYSIS
OF DISTRIBUTION DEMAND VARIATIONS

Metz Reference Room
University of Illinois
E106 NCEL
208 N. Romine Street
Urbana, Illinois 61801

by
GORDON GRACIE

Supported by
DIVISION OF WATER SUPPLY AND POLLUTION CONTROL
U. S. PUBLIC HEALTH SERVICE
RESEARCH GRANT WP-526

DEPARTMENT OF CIVIL ENGINEERING
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS
JUNE, 1965
AN ANALYSIS OF DISTRIBUTION 
DEMAND VARIATIONS

by

Gordon Gracie 
Assistant Professor 
of Civil Engineering

A Report on Part IV of an Investigation 
Supported by 
Division of Water Supply and Pollution Control 
U.S. Public Health Service 
Research Grant WP-526

Conducted by 
The Department of Civil Engineering 
University of Illinois 
Urbana, Illinois

June, 1965
PREFACE

The studies reported herein comprise Part IV of an investigation of "criteria for analysis of water distribution systems". The investigation has been supported by the Division of Water Supply and Pollution Control, U.S. Public Health Service, under Research Grant WP-526. Professor M. B. McPherson is principal investigator.


Work has been initiated on Part V, "Ground-Storage Booster Pumping Hydraulics".
An Analysis of Distribution Demand Variations

by

Gordon Gracie

Assistant Professor of Civil Engineering

Civil Engineering Department

University of Illinois

Urbana, Illinois

A paper to be presented at the 85th Annual Conference of the American Water Works Association, Portland, Oregon, Session 19, July 1, 1965, as part of a Joint Discussion on "Computer Studies for Distribution Design and Operation."
INTRODUCTION

The feasibility of simulating operation for design of a water distribution system using a digital computer is demonstrated in the companion paper by M. B. McPherson and R. Prasad. [1] In seeking optimum operating requirements for the future, it is necessary to use detailed demand projections extending over a number of years. As a prelude to synthesizing demand data, it is good practice to investigate the statistical variability in the original data. Any regularity that is discovered in the variability of the original data might then provide insight into preferred methods of demand synthesis.

The objective of this paper is to document a study of the daily demand variability for the water distribution systems of Champaign-Urbana, Illinois; Belmont High Service District, Philadelphia, Pennsylvania; and Denver, Colorado.

Through the excellent cooperation of the water works officials of these three systems, records of the hourly demands were obtained for a two-year period (1962-63) for both Champaign-Urbana and Belmont H.S., and records of the daily demands were obtained for a ten-year period (1954-1963) for Denver. For this study it was decided to investigate only the variability of daily demands. Thus, the hourly demands provided for Champaign-Urbana and Belmont H.S. were first processed in order to obtain average daily demand values. The data then available for study consisted of fourteen years of daily demands: two years of data for Champaign-Urbana, two years for Belmont H.S., and ten years for Denver.

*(References are listed at end of text).*
Champaign-Urbana and Belmont H.S. consume approximately the same amount of water, about 10 mgd on the average. Denver, on the other hand, consumes an average of well over 100 mgd. Thus the records of two relatively small water distribution systems of approximately the same annual demand and one large water distribution system were made available for study. All data were reduced to Standard Time. The raw data were used as received, without any adjustments for special extremes such as fire demands and outages.

Each of the fourteen years of daily demand data was subjected to the same statistical analysis. The 366th day of a leap year was purposely deleted from the corresponding Denver record in order to preserve consistency in the record lengths.

For the most part, only the results of the statistical analyses for the years 1962 and 1963 will be presented in this paper. This means that the statistical results for the eight years of Denver data, 1954 to 1961, will not be displayed. These statistical data are purposely omitted because it was felt their inclusion would only confuse their graphical portrayal rather than contribute anything to it, particularly when the graphs are superimposed upon each other. All ten years of Denver data will, however, be considered in the discussion of the year-to-year variation later on in the paper. It might be pointed out that the inclusion of two years of statistical results, rather than just one years worth, was made in order to indicate repeatability in the daily water demand from one year to the next for a particular system.
DATA PROCESSING

In general, several avenues of approach may be taken to analyze data in a statistical fashion. The objective in such analysis, of course, is to obtain those statistics which are meaningful and which, therefore, can be called upon to give an adequate description of the process from which the data are extracted.

As many statistical techniques were applied to the data as time would permit. Because there is an almost endless number of ways of manipulating the data, no attempt was made to exhaust all possibilities. However, the data were subjected to several well-established statistical operations, such as sample average and standard deviation evaluation, analysis of variance, harmonic analysis, autocorrelation coefficient evaluation, and spectral analysis.

Needless to say, the application of such techniques to fourteen years worth of daily demand data involves a tremendous number of computations. As a consequence, it would be virtually impossible to accomplish the work without utilizing a high-speed digital computer. The IBM 7094-1401 data processing system of the University of Illinois was used to execute the various computation operations required for the statistical processes.

Except for the harmonic analysis, all data processing was facilitated by using a special package of programs, called SSUPAC, \( ^{[2]} \) prepared by the Statistical Service Unit of the University of Illinois and residing permanently in the Disk Storage File of the computer system. The harmonic analysis was executed by using a special program located in the subroutine library of the computer's master system. Programs of this type are in fairly common usage throughout the country.
Much more statistical information was provided by the computer programs than was used for the investigation described in this paper. In order to give some idea of the magnitude of the amount of information produced by the computer, over 3000 sheets of output were generated in the data reduction process.

**DAILY AND WEEKLY AVERAGES**

Before considering the more involved techniques of harmonic analysis, autocorrelation coefficient evaluation and spectral analysis, it would be appropriate to describe the variation in the daily demands by using simple arithmetic averages. The daily demand data were first manipulated into a rectangular array consisting of 7 columns (one column for each day of the week) and 52 rows (one row for each week of the year). The 365th day of the year necessarily was not included in the array. The daily demand for each day of the week, averaged over the whole year, and the average daily demand for each week of the year were then computed.

The day-to-day variations in the average daily demand over all 52 weeks are given in Figure 1. For Champaign-Urbana it is apparent from Figure 1 that although the level of demand increases from 1962 to 1963, the shape of the average demand curve remains the same. It would appear that a fairly constant demand is maintained throughout the weekdays, whereas there is a lowering of demand on the weekends. For Belmont H.S. there is practically no variation at all from one day to the next. However, the change in the demand level from 1962 to 1963 is essentially the same (about 1 mgd) for Belmont H.S. as that for Champaign-Urbana. For Denver, there is virtually no change in the day-to-day demand level from 1962 to 1963, except for Friday and Saturday. For both years
FIG. 1 DAY-TO-DAY VARIATION
there is a distinct decrease in the demand on Sunday with respect to the other days. The records of the other eight years for Denver have a similar decrease.

The week-to-week variations in the average daily demand over the year are given in Figure 2. It is evident that there is less regularity in the week-to-week variations than there is in the day-to-day variations. This might well be expected because the day-to-day variations reflect essentially the weekly cycle of habits of a community, whereas the week-to-week variations arise from external conditions, such as changes in air temperature, which can be very irregular. It should also be remembered that the values in Figure 1 represent averages of 52 daily demands, while the values in Figure 2 represent averages of only 7 daily demands. The larger sample size will tend to produce smaller variation in the sample average. Month-to-month curves would undoubtedly exhibit greater regularity than the week-to-week curves.

If anything of significance can be interpreted from Figure 2, it is the rather overwhelming evidence that Denver uses much more water in the summer than it does in the winter compared with the other two systems. The water demand between the 46th week (mid-November) of one year and the 12th week (late March) of the following year remains unusually constant and at a very low level. Beginning at the 12th week, there is a distinct upward trend in the water demand until about the 26th or 27th week (early July) when the demand is at least three times that of wintertime. From the 27th week to the 46th week, the demand declines to the winter level. This characteristic of the Denver variation is typical of the other eight years.
FIG. 2 WEEK-TO-WEEK VARIATION
For Champaign-Urbana and Belmont H.S. there is a higher demand in the summertime than in the wintertime, although the change is not nearly as dramatic as that for Denver. The unusually low values for the 24th, 25th, and 26th weeks of the 1962 record of Belmont H.S. would seem to be abnormal, and may be nothing more than the result of a data transmission-logging malfunction.

Whether or not there is any statistical significance to the day-to-day and week-to-week variations portrayed in Figures 1 and 2 bears consideration. In order to test the statistical significance, the technique of analysis of variance [3][4] was employed. Variance is simply a measure of the deviation of the observed value from the mean. Analysis of variance breaks down the total variance of a sample of data into a number of component variances with the purpose of indicating how different sources of variability contribute to the overall variability of the data.

Three sources of variability were considered: the location of the particular day within the week, the location of the particular week within the year, and a random error component. The total demand \(Y_{ij}\) for a particular day \(i\) in a particular week \(j\) of the year may then be regarded as the sum of four components:

\[Y_{ij} = \bar{Y} + a_i + b_j + e_{ij}\]

where \(\bar{Y}\) = overall average demand for the entire year

\[a_i = \text{deviation from } \bar{Y} \text{ of the average demand for the } i\text{th day of the week}\]

\[b_j = \text{deviation from } \bar{Y} \text{ of the average demand for the } j\text{th week of the year}\]

\[e_{ij} = \text{random error component for the } i\text{th day in the } j\text{th week.}\]
There may exist some doubt as to the wisdom of applying an analysis of variance here in view of the fact that the data are not statistically independent. Therefore, the statistics so obtained should be interpreted cautiously.

The first objective of the analysis of variance is to test whether or not \( a_i \) and \( b_j \) are each significantly different from zero for each year of demand data. Pertinent results of the analysis are presented in Table 1. The significance of a particular variation is measured by its F ratio with due consideration for the associated number of degrees of freedom. Limiting values for the F ratio can be taken from any statistical table of the F statistic. Testing was at the 1% level of significance. For the variation between the days of the week, the theoretical upper limit is \( F = 2.87 \) (for 6 degrees of freedom in numerator and 306 degrees of freedom in denominator). For the variation between the weeks of the year, the corresponding theoretical upper limit is \( F = 1.60 \) (for 51 degrees of freedom in numerator and 306 degrees of freedom in the denominator). Any F ratio greater than these limiting values indicates statistical significance in the variation being tested.

The F ratios in Table 1 indicate significance in all variations except the variation in the Belmont H.S. demands between the days of the week. In other words, consideration of day-to-day variation in the Belmont H.S. system is not warranted. This lack of variation, of course, was already evident in Figure 1.

**HARMONIC ANALYSIS**

A list of values of a variate according to time is known as a time series. Each record of daily demand data is, therefore, a time series. It is assumed
TABLE 1

ANALYSIS OF VARIANCE

<table>
<thead>
<tr>
<th>System</th>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>1962 Mean Squares (mgd)^2</th>
<th>F Ratio</th>
<th>1963 Mean Squares (mgd)^2</th>
<th>F Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Champaign-Urbana</td>
<td>Between Days of Week</td>
<td>6</td>
<td>19.583</td>
<td>76.28</td>
<td>17.312</td>
<td>26.76</td>
</tr>
<tr>
<td></td>
<td>Between Weeks</td>
<td>51</td>
<td>7.218</td>
<td>28.11</td>
<td>17.053</td>
<td>26.36</td>
</tr>
<tr>
<td></td>
<td>Random Error</td>
<td>306</td>
<td>0.257</td>
<td>0.647</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belmont H. S.</td>
<td>Between Days of Week</td>
<td>6</td>
<td>0.682</td>
<td>0.96</td>
<td>0.577</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Between Weeks</td>
<td>51</td>
<td>11.912</td>
<td>16.78</td>
<td>8.098</td>
<td>12.16</td>
</tr>
<tr>
<td></td>
<td>Random Error</td>
<td>306</td>
<td>0.710</td>
<td></td>
<td>0.666</td>
<td></td>
</tr>
<tr>
<td>Denver</td>
<td>Between Days of Week</td>
<td>6</td>
<td>3012</td>
<td>3.00</td>
<td>4264</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>Between Weeks</td>
<td>51</td>
<td>33581</td>
<td>33.48</td>
<td>35259</td>
<td>28.77</td>
</tr>
<tr>
<td></td>
<td>Random Error</td>
<td>306</td>
<td>1003</td>
<td></td>
<td>1226</td>
<td></td>
</tr>
</tbody>
</table>
that each time series investigated in this study is stationary; that is, the statistical properties remain constant throughout the entire series involved.

An important item to be considered in the analysis of a time series is a search for periodicity. Periodicity may be simple; that is to say, it may be mathematically described by a single sine curve. If the period of this sine curve equals the total period investigated (one year, in the cases at hand), then this curve comprises what is known as the fundamental or first harmonic. Generally, more harmonics than the first are required to describe periodicity. The second harmonic has a period equal to one-half the fundamental, the third harmonic has a period equal to one-third the fundamental, and so on. At most, \( N/2 \) harmonics are required to generate a curve that will pass through \( N \) discrete points. Thus, for a time series consisting of 365 daily demands, at most 182 harmonics are required to satisfy the data.

Each daily demand value \( Y_t, \ t = 1, 2, \ldots, 365 \) can then be given by the following full series:

\[
Y_t = \bar{Y} + \sum_{k=1}^{182} (A_k \cos \omega tk + B_k \sin \omega tk)
\]

where \( \bar{Y} = \) overall average daily demand for the year

\( k = \) harmonic number

\( \omega = \frac{2\pi}{365} \) radians

\( A_k \) and \( B_k \) are constants associated with the \( k \)th harmonic.

If \( C_k = \) amplitude of the \( k \)th harmonic, and \( D_k = \) phase angle (point of change from a positive value to a negative value) of the \( k \)th harmonic in radians, it can be shown that
\[ C_k^2 = A_k^2 + B_k^2 \]

and \[ \tan D_k = \frac{A_k}{B_k} \].

\[ Y_t = \bar{Y} + \sum_{k=1}^{182} (C_k \sin (\omega t k + D_k)). \]

The determination of the constants \( A_k \) and \( B_k \), and hence \( C_k \) and \( D_k \), for a given time series is known as harmonic analysis. It can be shown that

\[ A_k = \frac{2}{365} \sum_{t=1}^{365} Y_t \cos \omega t k \]

and

\[ B_k = \frac{2}{365} \sum_{t=1}^{365} Y_t \sin \omega t k. \]

In addition to evaluating \( A_k \), \( B_k \), \( C_k \), and \( D_k \), the digital computer was programmed to determine the residual error for each day in the time series after the introduction of each successive harmonic. The root mean square residual (the square root of the mean of the squares of the residuals) was also determined after the introduction of each harmonic. After the 182nd harmonic is introduced, all residual errors must be zero and the data processing terminates.

In Figure 3, the amplitudes \( C_k \) of all 182 harmonics are plotted for 1962 and 1963, Champaign-Urbana.

For both years the amplitude of the first harmonic is the outstanding one, indicating the annual cycle of variation. The amplitude of the third harmonic is also quite large for 1963, but not for 1962. Two other harmonics of sizeable
FIG. 3 AMPLITUDES OF HARMONICS FOR CHAMPAIGN-URBANA
amplitude for both years are the 11th and 13th. However, the most interesting harmonics are the 52nd, 104th, and 156th. For both 1962 and 1963, these latter three harmonics rise well above the general level. This reflects a distinct weekly cycle in the records; the 52nd harmonic has a one-week period, the 104th harmonic has a one-half-week period, and the 156th harmonic has a one-third-week period. This corroborates the large F ratios in Table I for the variation between the days of the week for Champaign-Urbana.

The amplitudes of all harmonics for each of the years 1962 and 1963 for Belmont H.S. are plotted in Figure 4. Little similarity between the two years is evident. As a matter of fact, the second harmonic of the 1963 record is the predominant one of the year, whereas the second harmonic of the 1962 record is quite inconspicuous. Somewhat interesting in the 1962 data are the persistently larger amplitudes of the eight odd harmonics in the first 16 harmonics over the eight even harmonics.

Figure 5 shows the amplitudes of all harmonics for Denver for 1962 and 1963. The striking feature about the Denver data is the unusually large value for the amplitude of the first harmonic. This harmonic is at least 3½ times larger than the next largest harmonic, and is about 20 times larger than the average order of magnitude of all other harmonics. This, of course, indicates a decided annual cycle in the Denver demand data, a cycle which has already been clearly displayed in Figure 2. The 26th harmonic for 1962 is relatively large, indicating for that year at least, a biweekly oscillating pattern. The saw tooth configuration for the summer portion of the 1962 Denver record in Figure 2 also illustrates this oscillating feature. Notice, however, that the winter portion of the 1962 Denver record is devoid of this oscillating feature. This means that there exists
FIG. 5 AMPLITUDE OF HARMONICS FOR DENVER

Amplitude of Harmonic Number 1 = 88.14 mgd
Number 4 = 25.95 mgd
Number 5 = 23.84 mgd

Amplitude of Harmonic Number 1 = 87.46 mgd
Number 2 = 21.60 mgd

1963

1962
another harmonic which is in phase with the 26th harmonic during the summertime, producing the saw tooth effect, but which is out of phase with the 26th harmonic during the wintertime when it produces a smooth record. The other harmonic must, by the very nature of the effect, be a neighbor of the 26th harmonic and have a phase angle about $\pi$ radians different from that of the 26th harmonic. It so happens that the 25th harmonic is also quite large and has a phase angle different from that of the 26th harmonic by 3.20 radians.

There is some evidence of a weekly cycle in the Denver data. The evidence is in the form of peaks at or next to the 52nd and 156th harmonics of the 1962 record, and at or next to the 52nd, 104th, and 156th harmonics of the 1963 record. (Much stronger evidence of a weekly cycle appears in the 1956 Denver record. For that record, not shown here, there are very evident peaks at or next to the 52nd, 104th, and 156th harmonics. Perhaps the weekly habits of the Denver community changed between 1956 and 1962. At least, the statistical data would seem to bear this out.)

To conclude this section, it might be interesting to investigate the improvement afforded by each harmonic in the fit of the mathematically generated curve to the original data. In Figures 6 and 7 are plotted the root mean square (square root of the mean of the squares) of the residuals after each successive harmonic has been introduced. Figure 6 displays the plotted data for Champaign-Urbana and Belmont H.S., and Figure 7 displays that for Denver.

The root mean square of the residuals necessarily must be a decreasing function with respect to the harmonic number. A sharp drop in the function indicates a significant improvement in the fit of the mathematical curve to the original data provided by the harmonic at which the drop takes place. This is
FIG. 6 RESIDUAL ERROR AFTER FITTING SUCCESSIVE HARMONICS
FIG. 7 RESIDUAL ERROR AFTER FITTING SUCCESSIVE HARMONICS
but another way of pointing out the significance of certain harmonics, in addition to drawing attention to the similarity between two different years' worth of data taken from the same system.

The curves in Figure 6 for Champaign-Urbana perhaps point out the similarity between different years most effectively. Indeed, if the 1963 values were reduced by 10%, the percentage by which the average 1963 daily demand exceeded its 1962 counterpart, there would be an unusually good correspondence between the two curves. The abrupt changes in each curve at the 52nd, 104th, and 156th harmonic again graphically reveal the existence of a weekly cycle.

The curves for Belmont H.S. display no particular significance in any harmonic beyond the 15th. Particularly in the 1963 record, it would seem that every harmonic in the entire series, except perhaps the 2nd, 3rd, and 5th harmonics, contributes an equal share in fitting the data. For 1962, the first 15 harmonics take care of over 50% of the variation. The lower starting point for the 1963 curve, particularly in view of the fact that the average 1963 daily demand exceeds the average 1962 daily demand by 10%, indicates that there is less variation in the 1963 data than there is in the 1962 data.

The curves for Denver in Figure 7 show once more the major contribution of the first harmonic. After fitting the first harmonic to the data, the root mean square of the residuals is reduced to about 60% of its initial value for 1963, and to about 56% of its initial value for 1962. The significant changes in the 1963 curve at the 4th and 5th harmonics indicate their important contributions in fitting the data; the same can be said for the 2nd harmonic of the 1962 data. The improvement made by the first five harmonics is almost identical
for both years. For 1962, the first five harmonics reduced the root mean square of the residuals to 50% of its initial value, and for 1963 to 49% of its initial value; that is, in both Denver cases, the first five harmonics take care of one-half the variation in the original data. Changes in the curves at the 52nd, 104th, and 156th harmonics again reflect some kind of weekly cycle in the daily demands.

AUTOCORRELATION AND SPECTRAL ANALYSIS

The graphs exhibited in Figures 3, 4, and 5 provide some idea of the contributions being made by the individual harmonics to the overall variation in the time series of demand data. In fact, a simple relationship exists between the amplitude of a harmonic and its contribution to the variance of the data. This relationship is: [4]

\[ V_k = \frac{C_k^2}{2} \]

where \( V_k \) = contribution of the kth harmonic to the variance
\( C_k \) = amplitude of the kth harmonic.

A plot of \( V_k \) versus k may then be considered as a sample of the spectrum of contributions of all harmonics, or frequencies, to the total variance of the time series. Unfortunately, when the \( V_k \) are plotted for all harmonics, the points scatter widely. A similar but less emphatic scatter pattern is given by the \( C_k \), as evidenced in Figures 3, 4, and 5. The scattering arises from the fact that the harmonics are being computed from a random sample of data, a sample that is subject to statistical fluctuation. What is really sought in the analysis of the times series is the underlying smooth spectrum of frequencies which describes the process producing the time series in the first place.
It is possible to estimate the underlying smooth spectrum by applying some algebraic smoothing process to the series of $V_k$ values. However, the preferred technique of estimating the smooth spectrum is that of spectral analysis.

The technique of spectral analysis is based upon autocorrelation functions. The autocorrelation of a time series means correlation of the time series with itself; that is, the correlation of a time series with the same time series an interval of time later. Thus, the autocorrelation of the time series $Y_t$, $t = 1, \ldots, 365$ means the correlation of the values $Y_t$ with those of $Y_{t+p}$, $t = 1, \ldots, 365-p$ where the quantity $p$ is known as the lag.

If the demand data are first normalized by deducting the mean and dividing by the standard deviation, then the autocorrelation coefficient, designated $r_p$, can be evaluated using the expression:

$$r_p = \frac{\sum(Z_t \cdot Z_{t+p})}{\Sigma(Z_t)^2} \quad t = 1, \ldots, N-p$$

where $Z_t$ is the normalized form of $Y_t$ and $N$ is the number of points in the series ($N = 365$ for the cases at hand).

Like the usual correlation coefficient, the autocorrelation coefficient ranges between +1 and -1. Naturally, the correlation coefficient for $p = 0$ must be +1. If the autocorrelation coefficient maintains a relatively high positive value as the lag increases, then the data possess the property of persistence; that is, a high demand one day leads to a high demand the next day or several days, or vice-versa. The coefficient may eventually become negative after a certain lag period. This means that the data is tending to balance out; that is, a high demand over one period of time will be followed by a period of low demand.
The plot of the autocorrelation coefficient $r_p$ with respect to the lag $p$ is called a correlogram. Figure 8 shows correlograms for the years 1962 and 1963 for Champaign-Urbana, Belmont H.S. and Denver.

Of the three systems, Champaign-Urbana has the most interesting correlograms. The 1962 correlogram takes the form of a gentle curve starting at 0.7 for $p = 0$ and decreasing to a constant level of 0.2 for all $p$ greater than 15, upon which are superposed peaks of height 0.3 at each interval of 7 lags, beginning with $p = 0$. The peaks themselves are almost identical in shape and are very symmetrical. The 7-lag interval is, of course, indicative of the weekly demand cycle; but the other regular features of the correlogram are difficult to explain.

The 1963 Champaign-Urbana correlogram takes on a form similar to the 1962 correlogram. The apexes of the peaks are about at the same level as those of the 1963 graph, but the base level of the curve is much higher and rapidly approaches zero at $p = 34$.

The Belmont H.S. correlograms appear to be quite different, so different in fact, that it is difficult to believe that they are derived from samples of presumably the same time series. It is likely that this disparity arises from the abnormally low demands for the 24th, 25th, and 26th weeks of 1962 (see Figure 2). This anomaly in the 1962 record is so pronounced that it could strongly affect the statistical information derived from it.

The Denver correlograms match relatively well. The 1963 correlogram is a reasonably smooth curve with a gentle change in slope at about $p = 4$. The 1962 curve is wavier than the 1963 curve, faintly indicating a weekly period.
FIG. 8  CORRELOGRAMS
Autocorrelation may be expressed not only in terms of the autocorrelation coefficient \( r_p \) but also in terms of the autocovariance function \( W_p \). The autocovariance function is defined by [5]

\[
W_p = \frac{1}{N-p} \sum_{t=1}^{N-p} (z_t \cdot z_{t+p})
\]

where the symbols used are as defined previously. It follows that

\[
r_p = \frac{W_p}{S^2}
\]

where \( S^2 \) is the sample variance.

The spectral density of a discrete valued time series can be estimated by applying a finite Fourier series transformation to the autocovariance function. In simpler terms, this means that an estimate of the spectral density is obtained by applying a harmonic analysis to the set of autocovariance values, \( W_0, W_1, \ldots, W_M \), where \( M \) is the maximum value of the lag \( p \). \( M \) is ordinarily chosen as not over 10% of \( N \). [5] \( M = 34 \) was selected.

Raw estimates \( L_p \) of the spectral density are obtained using the following expression

\[
L_p = W_0 + 2 \sum_{q=1}^{M-1} W_q \cos \frac{q \pi p}{M} + W_M \cos \pi p.
\]

Smoothing of these raw values is accomplished by using the following weighted moving average

\[
U_p = 0.23 L_{p-1} + 0.54 L_p + 0.23 L_{p+1}
\]

(where \( L_{-1} = L_1 \) and \( L_{M+1} = L_{M-1} \)).
For a more detailed discussion of the above expressions see [5]. Thorough discussion of spectral analysis may be found in [6] and [7].

The smooth spectrum is then formed by plotting the $U_p$ values and passing a smooth curve through them by eye. Figure 9 shows the smooth spectra for Champaign-Urbana, Belmont H.S., and Denver formed in this manner. The area under each curve within a prescribed interval is to be interpreted as the contribution of the frequencies in that interval to the overall variance of the time series. All values $U_0$ and $U_1$ are extremely high, being well off the plotting range. This feature emphasizes the major contribution of the low frequency harmonics to the overall variation of the data.

In Figure 9 the weekly cycle of the Champaign-Urbana data is again well defined. The weekly periodicity centers about 9.7 cycles/68 days (1 cycle/7 days), the semi-weekly periodicity centers about 19.4 cycles/68 days (2 cycles/7 days), and the three-times weekly periodicity centers about 29.1 cycles/68 days (3 cycles/7 days).

The Belmont H.S. spectra do not coincide very well, particularly around 7 cycles/68 days. This lack of agreement is perhaps another manifestation of the large anomaly in the original 1962 data.

The Denver spectra also do not agree as well as the Champaign-Urbana spectra. Nevertheless, a weekly cycle is mildly revealed.

In setting up a model for simulation purposes, the spectral density function is probably one of the best things to use. It can be employed to select
FIG. 9 SPECTRAL DENSITIES
those frequencies which contribute most to the variation in the distribution demand and which, therefore, generate an authentic representation of the underlying process.

YEAR-TO-YEAR VARIATIONS

The ten years of Denver data provide a substantial quantity of information which can be put to good use in analysing the variation in the distribution demand of a system from year to year. Certain parameters, or more correctly, the statistical estimates of these parameters, were selected and plotted on a yearly basis for the ten years 1954 to 1963. The parameters selected were the average daily demand for each year, the standard deviation of the data for each year, the amplitudes of the first and second harmonics, and the phase angles of the first and second harmonics.

Figure 10(a) is a plot of the average daily demand for each year. It is obvious that there is a decided upward trend in the water demand for Denver with respect to time. Figure 10(b) is a plot of the standard deviation for each year. This too has a decided upward trend, which leads one to believe that a higher degree of variation is associated with a higher level of demand. The ratios of the standard deviations to their corresponding average daily demands are plotted in Figure 10(c). A small upward trend appears to exist in these ratios; however, if a linear regression line were fitted to these plotted data, it is doubtful that the slope of the line would be statistically significant. A ratio of about 0.5 does, of course, indicate an extremely high degree of variation in the distribution demand, confirming previous evidence for the Denver data.
FIG. 10 YEAR-TO-YEAR VARIATION FOR DENVER
Figure 10(d) displays the variation in the amplitudes of the first and second harmonics from year to year. Because the first harmonic is by far the dominant one for Denver, it is to be expected that in addition to being very large its amplitude will increase as the level of variation increases. Therefore, it exhibits a pattern very similar to that for the standard deviation. The amplitude of the second harmonic, on the other hand, is much smaller than that of the first harmonic, and its pattern does not resemble that of the standard deviation at all. In Figure 10(e) are plotted the ratios of the amplitudes of the first and second harmonics to the corresponding average daily demands respectively.

Figure 10(f) shows the variation in the phase angles of the first and second harmonics from year to year. The significant feature displayed by this graph is the constancy of the phase angle for the first harmonic. This means that the zero point of the first harmonic occurs at just about the same time of the year, year after year, at about the 37th week. Indeed, if all ten years of the Denver demand data were strung out into one long 10-year series, it should be possible to obtain a remarkable fit to these data with a simple mathematical curve of the form:

\[ y_t = \bar{v} + C_1 \sin \left( \frac{2\pi t}{365} + 4.5 \right) \]

where \( t \) is the number of the day in the 10-year period, and \( \bar{v} \) (the average daily demand) and \( C_1 \) (the amplitude of the first harmonic) are linear functions of \( t \). This model differs considerably from that used for natural hydrologic data\,[8], such as for stream flow, where future conditions are projected on the assumption that the process is completely random without trend.
CONCLUDING REMARKS

A number of statistical techniques were used to analyze the variability in the daily demands of three water distribution systems with the objective of disclosing regularity in the demand data. The water distribution systems studied were Champaign-Urbana, Illinois; Belmont High Service District, Philadelphia, Pennsylvania; and Denver, Colorado. The statistical evidence was displayed in various forms such as graphs of simple averages, harmonic amplitudes, correlograms, and spectral densities.

Definite annual and weekly cycles were found to exist. The most pronounced annual cycle was that for Denver, and the most pronounced weekly cycle was that for Champaign-Urbana. Other features were also exhibited.

By no means does this study exhaust all avenues of investigation of the observed data. It is very likely that other items of regularity could be discovered using other techniques. However, it is felt that there is substantial evidence from this investigation to justify consideration of deterministic content when formulating a mathematical model of the distribution demand of a system. Certain features of regularity seem to be associated with a particular distribution system. Indeed, a simulation model which ignores these features would probably be inadequate to describe the underlying process producing the demand data.
ACKNOWLEDGEMENTS

This study covers a portion of an investigation entitled "Criteria for Analysis of Water Distribution Systems," supported by the Division of Water Supply and Pollution Control, U.S. Public Health Service, Research Grant WP-526. Professor M. B. McPherson is principal investigator.

The two years of Belmont H.S. District hourly demand data were supplied through the courtesy of Commissioner S. S. Baxter, by Mr. J. V. Radziul, R & D Unit Chief, Philadelphia Water Department. Copies of Load Control Center log sheets were provided, from which cards were punched directly.

The two years of Champaign-Urbana hourly demand data were obtained through the courtesy of Mr. E. R. Healy, Vice Pres. and Mgr., and Mr. R. S. Shierry, Distribution Superintendent, Northern Illinois Water Corp., Champaign, Illinois. Hourly rates from operating logs of the two treatment plants and hourly records of water levels for the three equalizing storage tanks were reduced to hourly demands by project personnel, Messrs. R. K. Rai and N. Chaudhari, and then punched on cards.

The ten years of Denver daily demand data were supplied through the courtesy of Mr. R. B. McRae, Chief Engineer, by Mr. R. C. McWhinnie, Hydraulic Engineer, Board of Water Commissioners, Denver, Colorado. Cards were punched directly from demand listings provided.
REFERENCES


