1971

Analysis and Prediction of Telephone Demand in Local Geographic Areas

Douglas M. Dunn

William (Bill) H. Williams
CUNY Hunter College

W. Allen Spivey

Analysis and prediction of telephone demand in local geographical areas

Douglas M. Dunn and William H. Williams

Members of Technical Staff
Bell Telephone Laboratories

and

W. Allen Spivey

Professor of Statistics
University of Michigan

An approach to forecasting the demand for local area telephone service is presented in this paper. The specific problem discussed is the forecasting of main stations in three Michigan metropolitan areas. Several different statistical models are used. The first class of models introduced uses adaptive exponential smoothing and is based solely on the past history of the time series involved. Although appropriate data at the local area level are very difficult to obtain, two exogenous time series related to household formations are used to construct more elaborate models for one of the areas. The various models are evaluated by both the average absolute and the root-mean-square forecast error. In terms of these criteria, the first class of models referred to above performs reasonably well while the second set does considerably better. This argues strongly that future improvements in forecasting accuracy will be made by the more extensive involvement of exogenous variables.

1. Introduction

■ Forecasting the demand for main telephone stations is a major management task in the Bell System which requires attention on many different fronts. Most of the usual problems in the development and interpretation of forecasts are present in this task. Moreover, it offers complex problems of its own which arise largely from

---

Douglas M. Dunn earned his B.S. and M.S. degrees at the Georgia Institute of Technology and the Ph.D. degree at the University of Michigan (1970). He formerly taught at the Georgia Institute of Technology (1965–1966) and began working at Bell Laboratories in 1970.

W. H. Williams received the A.B. degree from McMaster University (1954) and the M.S. and Ph.D. degrees from Iowa State University (1956 and 1957, respectively). He has held faculty appointments at McMaster University, Iowa State University, and the University of Michigan, and has been a consultant to the U. S. Census Bureau and the Executive Office of the President. Dr. Williams has published numerous papers in various statistical and economic journals. He has been employed at Bell Laboratories since 1960.

W. Allen Spivey earned the A.B., M.A. (1952), and Ph.D. (1956) degrees at the University of North Carolina. He has held positions at Kyoto University (Japan), the London School of Economics, and Harvard University and has served as an Associate Editor of *Management Science*. Dr. Spivey is the author or coauthor of several books and has published widely in various technical journals.

The authors are indebted to Michigan Bell Telephone Company for their support of this study. This paper is drawn from Dr. Dunn's Ph.D. thesis and associated work done both at Bell Telephone Laboratories and the University of Michigan.

the specific geographical context for which the forecasts must be prepared. For example, in forecasting at the level of the national economy or of the total Bell System, the geographical context may not be an important factor.[1] As the geographical area becomes smaller, however, one must take account of this influence. Exogenous data become much harder to obtain, and short-term local swings caused by strikes, welfare policy, and changing deposit practices appear to be impossible to predict by analytic techniques. It is interesting to note that forecasting methodologies used at the national level are often applied to a local area without any modification whatsoever.

An additional concern in this problem is the need to forecast the demand for different classes of Bell System services. Two of the most important service classifications, business and residence main stations, were considered in this work. Although many other user services exist (for example, rural, coin, or mobile phones), they will not be considered in this paper.

The Bell System makes very large and growing annual expenditures in construction. Recently, this amount was in the neighborhood of seven billion dollars, with an annual rate of growth of approximately 10 percent. Much of this investment goes toward the construction or expansion of wire centers[2] and associated outside plant equipment in order to tie the telephone user to the switching equipment. Clearly, local area forecasts coupled with additional information are an important input into this construction program. The magnitude of the forecasting problem is seen when one considers that for the Michigan Bell Telephone Company, the operating company supplying the local area data used in this study, there are over 300 wire centers for which forecasts are needed on at least an annual basis.

☐ **Local area forecasting.** Ideally, the forecasting activity at the local level should be closely related to the study of the complex interface between the forecasting of main stations and the planning and capital budgeting functions. For example, new facilities planning should be executed so that new equipment will become operational in each wire center just as the maximum capacity of the wire center is reached. Such a good match of forecasting and planning would allow new demand to be met, while at the same time capital costs of the increment would be held to a minimum.

In addition, the smaller the geographical area studied the more difficult is the problem of data collection, since far fewer data are available for counties and standard metropolitan areas than for states and still larger areas. Worse yet, forecasts should be made at the wire center level (i.e., the geographical area served by a given wire center), and this may be even smaller than a single county. No less frustrating is the fact that some wire centers include parts of two different counties.

Local area main station forecasts are needed for a period of approximately twelve to eighteen months in the future because this is the period of time required to design, manufacture, and install an addition to an existing wire center. Moreover, the engineering of

[1] See, for example, Chaddha and Chitgopekar [2]. For a more general discussion, see Box and Jenkins [1].

[2] A wire center may be defined as the physical facility housing the telephone switching equipment for a given geographical area.

additional facilities is of such a nature that it may be either impossible or prohibitively costly to install the additional equipment in another wire center should it turn out that demand for service by a given wire center does not grow as anticipated.

Although a forecast interval of eighteen months may be regarded by some as a short term for which forecasts should be relatively easy to develop, it is sufficiently long for the problem to be difficult. In addition, decisions based on these forecasts involve large amounts of money, as was indicated earlier.

In this work it has been assumed that the historical time series of main stations adequately represents the demand for main stations in the past. The validity of this assumption will vary from area to area. It is probably a good assumption if an area has few held orders.[3] On the other hand, if held orders exist they clearly represent part of the main station demand. A compounding difficulty is that in areas in which it is well known that telephone orders are not being filled, potential customers may not even request service, so that demand may exist which is completely unknown to the telephone company. In the areas used in this study, held orders do not appear to be a serious concern.

□ **The history of the study.** The work set forth in this paper had its origin in an association between staff members of Bell Telephone Laboratories and persons affiliated with Michigan Bell Telephone Company who were interested in refining and improving main station forecasts for local areas. After some experimentation with multiple-regression models (whose results were poor in terms of forecast accuracy), historical data of monthly main stations were studied using the techniques of statistical time series analysis, such as exponential smoothing, autoregression, and spectrum analysis. These procedures were used largely because good historical data on main stations were available on a monthly basis and data on related exogenous variables were scarce. Effort was concentrated on these models despite the authors' suspicion that future demand for telephones in a local area might be only weakly related to characteristics of historic demand. For example, main stations as of a given month may be more heavily influenced by zoning changes and shifting population mobility characteristics than by the past data on main stations in the area.

□ **The data.** The investigation began with two time series—a series of business main telephones and a series of residence main telephones—from each of three Michigan cities. Observations were available monthly for a fifteen-year period, from 1954 to 1969. The cities were selected to reflect different economic and demographic influences that must be dealt with successfully if one is to develop local area forecasting models of general usefulness. One of the cities, Flint, is large and depends heavily on one cyclical industry and its related firms; a second, Grand Rapids, is a city of about one-quarter of a million people and has a well-diversified economic base; and the third, Battle Creek, is a smaller city with a stable economic growth record.

Early in the study, the conclusion was reached that the business and residence series were sufficiently different to warrant separate

[3] A "held order" is a customer's request for a particular type of service which cannot currently be supplied.

analysis; only the analysis of residence telephones is presented in this paper. Furthermore, each series was decomposed into a series of connects and a series of disconnects, the difference of which yields the net main station gain for each month. Most of the analysis was directed towards these connect and disconnect series. There were five reasons for this:

(1) The initial analysis of aggregated series of residence main stations did not yield sufficiently accurate forecasts. Consequently, it was felt that the additional parameters which are introduced by the fitting of subaggregate models might yield lower forecast errors.

(2) It was evident that the series of connects and disconnects had different characteristics. As an extreme example of this difference, consider a resort area wire center in which connects and disconnects are stable all winter but in which the connect series regularly peaks in the spring and the disconnects regularly peak in the fall.

(3) The decomposition of main stations into connects and disconnects was feasible, since the operating company had the necessary data.

(4) The forecasting techniques developed for connect or disconnect series might be useful in forecasting other local area time series.

(5) Forecasts generated for the connect and disconnect time series are useful to the local operating companies in scheduling related activities. For example, the work of installers is closely tied to the levels of connect and disconnect activity.

2. Forecasting connect and disconnect series

■ Plots of the residence main station, connect, and disconnect series for Flint are shown in Figures 1 and 2. The most obvious characteristic of the main station series shown in Figure 1 is its volatility, a substantial portion of which reflects movements in the United States economy. For example, the large drops shown in the plot for the years 1958–1959 and 1961–1962 occur during periods of softening in the economy. There is a marked change in the rate of growth in the series beginning in 1966, and the very rapid advance in the number of residence telephones in service shown from 1962 to 1966 is not present in the years 1966–1969.

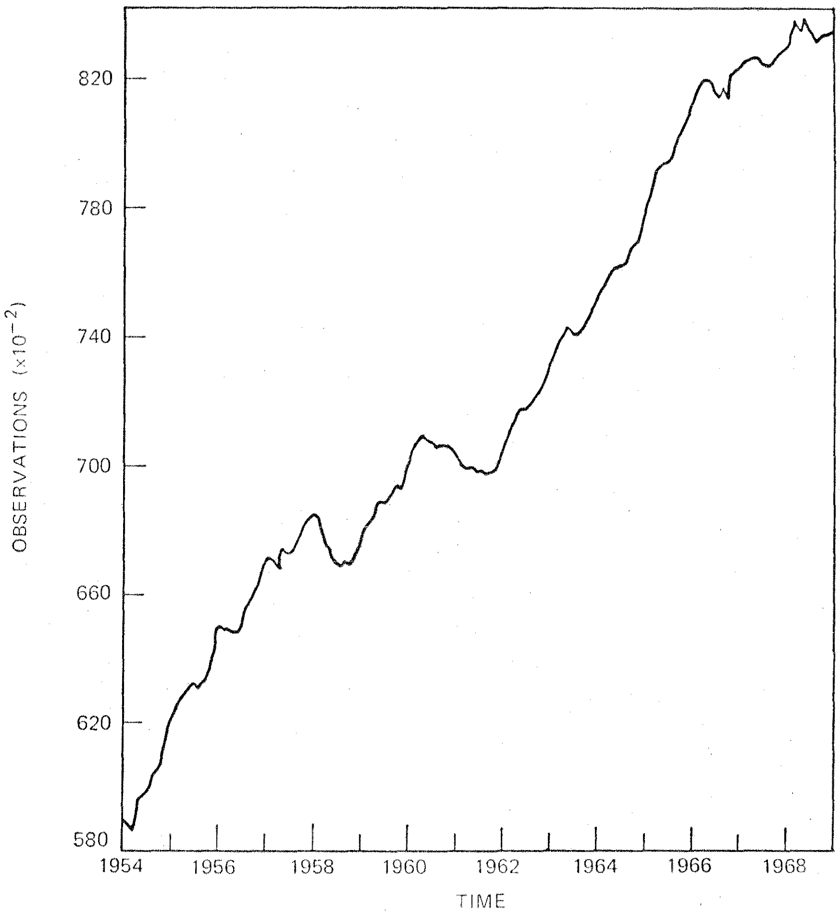
The series of connects and disconnects shown in Figure 2 exhibit a much stronger seasonal pattern than that seen in the main station series. This characteristic can be easily related to traditional moving patterns of families in the United States. Family moves occur with high frequency during the summer, and activity in connects and disconnects increases greatly in resort areas during the same period. Generally, the connect series is more volatile and less stable than the disconnect series.[4] Notice in Figure 2 that, for fixed periods of time, the connect series shows more peaks than does the disconnect series during the corresponding time period. Also note that the seasonal pattern tends to be more regular in the disconnect series than in the connect series.

☐ **Initial analysis.** Three major results were obtained from the initial analysis. First, the long-term behavior for the majority of the series

[4] In each of the local areas investigated, the connect series showed more power in the high frequency region of the spectrum than did the disconnect series.

FIGURE 1

RESIDENCE MAIN STATIONS FOR FLINT, MICHIGAN



is modeled well by a quadratic polynomial. Second, the residuals from the estimated quadratic polynomial are strongly seasonal in the sense that the power spectrum of the residuals shows a significant peak at a frequency corresponding to a twelve-month cycle. Third, the analysis (including the cross-spectrum) revealed no stable lagged relationships between the connect and disconnect series, so that subsequent analysis was directed toward forecasting each series separately and not toward forming an interactive connect-disconnect model.

Consequently, the first model was developed to forecast either the connect or disconnect series separately, and it assumed that the series consisted of a trend component $tr(t)$, a seasonal component $s(t)$, and an error term $\epsilon(t)$ in the additive manner displayed in equation (1). In equation (1), $\epsilon(t)$ is an independent random error with zero mean and finite variance σ_ϵ^2 for all t, and $y(t)$ may denote either the connect or the disconnect series:

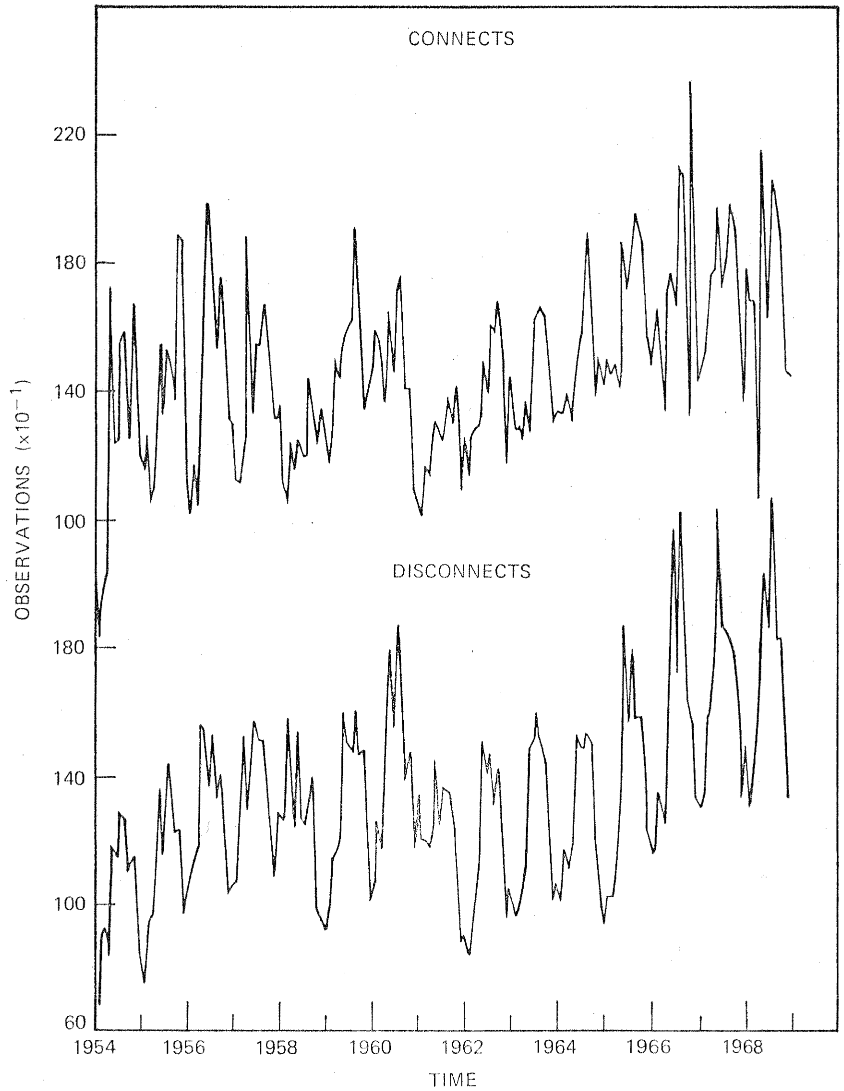
$$y(t) = tr(t) + s(t) + \epsilon(t); \quad t \in T. \tag{1}$$

The trend effect was forecast using a quadratic polynomial whose coefficients were estimated by standard least-squares regression techniques. These estimates were based on all data available at the time the forecast was made. When the forecast evaluation procedures described later in this section were followed, iterative methods[5] were

[5] See [3].

FIGURE 2

RESIDENCE CONNECTS AND DISCONNECTS FOR FLINT, MICHIGAN



used to update the estimated coefficients as new observations were incorporated. Thus, for time t, the estimated trend was

$$\hat{tr}(t) = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2 \,, \tag{2}$$

and the residual series,

$$r(t) = y(t) - \hat{tr}(t) \tag{3}$$

(an estimate of the seasonal and random components), was forecast using simple exponential smoothing. However, because of the strongly seasonal nature of the residual series, a decomposition was performed resulting in twelve separate monthly series: a series of January values (one from each year), a series of February values, and so forth for each month of the year. Simple exponential smoothing was then used on each of the twelve monthly series. This model was not satisfactory, and later sections describe further modifications.

☐ **Evaluating forecasting models.** In order to achieve effective measures of forecast results, one would like to have a loss function associated with the forecast errors for each local area. However, such loss functions are not easily determined. It is very difficult to associate dollar costs with unfilled demand for services or with excess plant capacity. Nevertheless, for this problem it appears to be reasonable to give the same weight to both positive and negative forecast errors of the same magnitude. Consequently, the average absolute forecast error and the root-mean-square forecast error were used for model evaluation. It is helpful to compute these measures by averaging over the time series in the following iterative way:

(1) Consider a monthly time series $y(t)$, $t \in T = \{1, 2, \ldots, N\}$, and use B months of data as a base period to obtain a forecast $\hat{y}(B + L)$, of the series for time $(B + L)$, where L is the lead time in months.

(2) Iterate through the series until data through period $(N - L)$ are used to forecast period N.

(3) Evaluate:

(a) Average absolute forecast error

$$\text{A.E.} = \frac{1}{(N - B - L + 1)} \sum_{k=B+L}^{N} |y(k) - \hat{y}(k)| .$$

(b) Root-mean-square forecast error

$$\text{R.E.} = \left[\frac{1}{(N - B - L + 1)} \sum_{k=B+L}^{N} [y(k) - \hat{y}(k)]^2\right]^{\frac{1}{2}} .$$

□ **Adaptive exponential smoothing.** As stated previously, the results obtained from use of a quadratic trend and a simple exponential smoothing of the residuals were not sufficiently accurate for our purpose. Consequently, adaptive exponential smoothing,[6] which is designed to allow the model to adapt to the dynamic changes in the trend component, was introduced. With this type of estimation, a given series, say $Z(t)$, is forecast as

$$\hat{Z}(t + L) = S_t(Z) , \tag{4}$$

where $S_t(Z)$ is defined in a manner similar to that for the smoothed function of the observations in simple exponential smoothing. Specifically,

$$S_t(Z) = \alpha(t)Z(t) + [1 - \alpha(t)]S_{t-1}(Z) . \tag{5}$$

The major difference between adaptive exponential smoothing and simple exponential smoothing is the dynamic assignment of the smoothing weight, $\alpha(t)$. To discuss this further, the following definitions are useful:

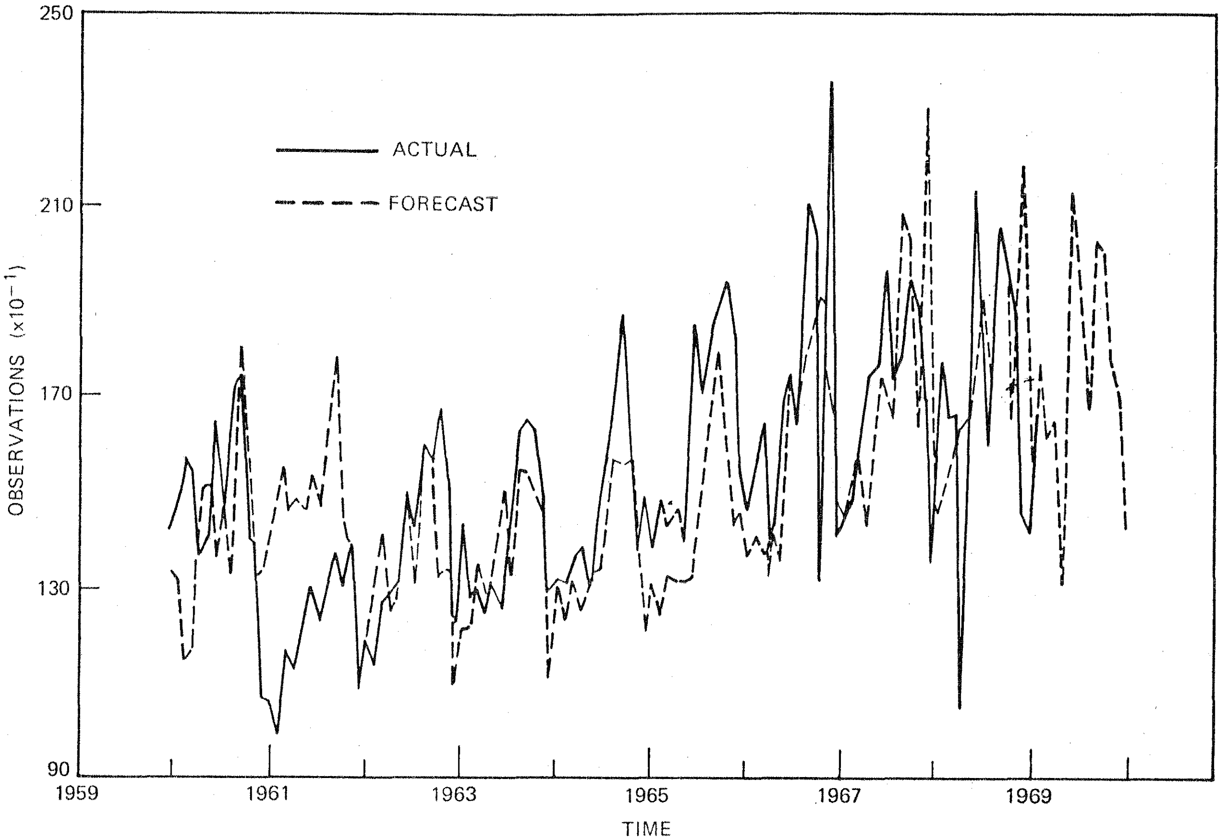
$$\alpha(t) = |\text{Tracking Signal } (t)| , \tag{6}$$

$$\text{Tracking Signal } (t) = \frac{SE(t)}{SAE(t)} , \tag{7}$$

[6] See [4].

FIGURE 3

FORECAST OF CONNECTS FOR FLINT, MICHIGAN



$$SE(t) = \gamma e(t) + (1 - \gamma)SE(t-1) \quad (8)$$

$$SAE(t) = \gamma |e(t)| + (1 - \gamma)SAE(t-1) \quad (9)$$

$$\Bigg\} \; 0 < \gamma < 1.0 ,$$

$$e(t) = [Z(t) - \hat{Z}(t)] . \quad (10)$$

The values of $SE(0)$ and $SAE(0)$ are defined to be zero, which yields $\alpha(1) = 1.0$, thus obviating the need to define $S_0(\hat{Z})$ in equation (5).

It is interesting to look at limiting values of the smoothing weight, $\alpha(t)$. Consider the case where all the forecast errors are of the same sign. In this case, the SE (smoothed error) and the SAE (smoothed

TABLE 1

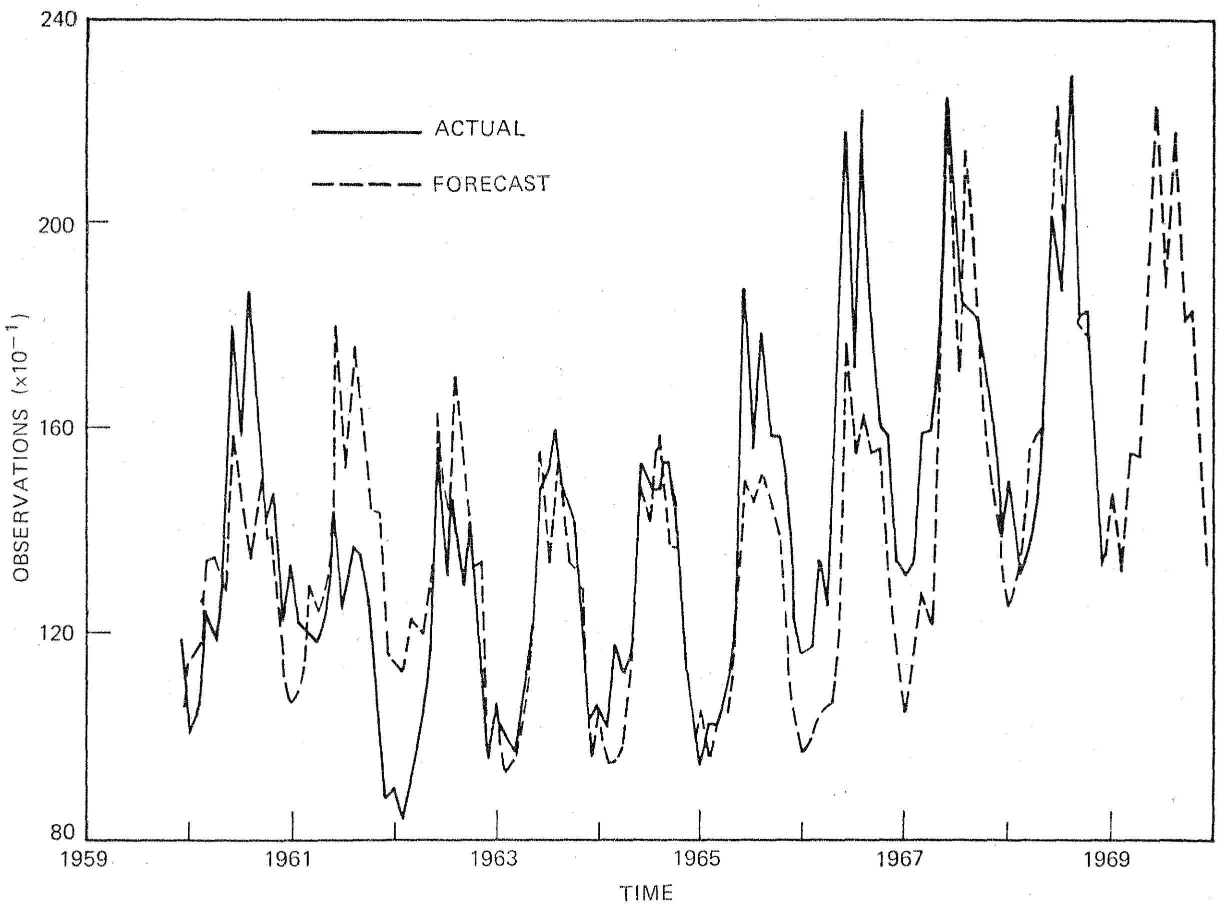
ADAPTIVE TWO-WAY FORECASTING MODEL, EQUATION (13)
(MAIN STATIONS)

LOCAL AREA	LEAD TIME	γ	β	FORECAST ERROR	
				AVERAGE ABSOLUTE	ROOT-MEAN-SQUARE
FLINT	12	0.9	0.8	1250.6	1588.1
	24	0.1	0.3	2124.0	2537.9
GRAND RAPIDS	12	0.7	0.4	562.5	723.0
	24	0.9	0.1	1034.3	1163.7
BATTLE CREEK	12	0.3	0.7	279.7	368.4
	24	0.3	0.6	724.2	842.7

NOTE:
ALL VALUES SHOWN IN THE BODY OF THE TABLE CORRESPOND TO THE LOWEST AVERAGE, ABSOLUTE FORECAST ERRORS OBSERVED FOR THE (γ, β) VALUES TESTED. IN EACH CASE, N = 180 MONTHS (SERIES LENGTH) AND B = 60 MONTHS (BASE PERIOD).

FIGURE 4

FORECAST OF DISCONNECTS FOR FLINT, MICHIGAN



absolute error) have the same magnitude. Hence the absolute value of their ratio is one, as is the value of $\alpha(t)$. A smoothing constant with value one gives the smoothed history of the series zero weight. Hence in this case very rapid adaptation to changes in the series occurs.

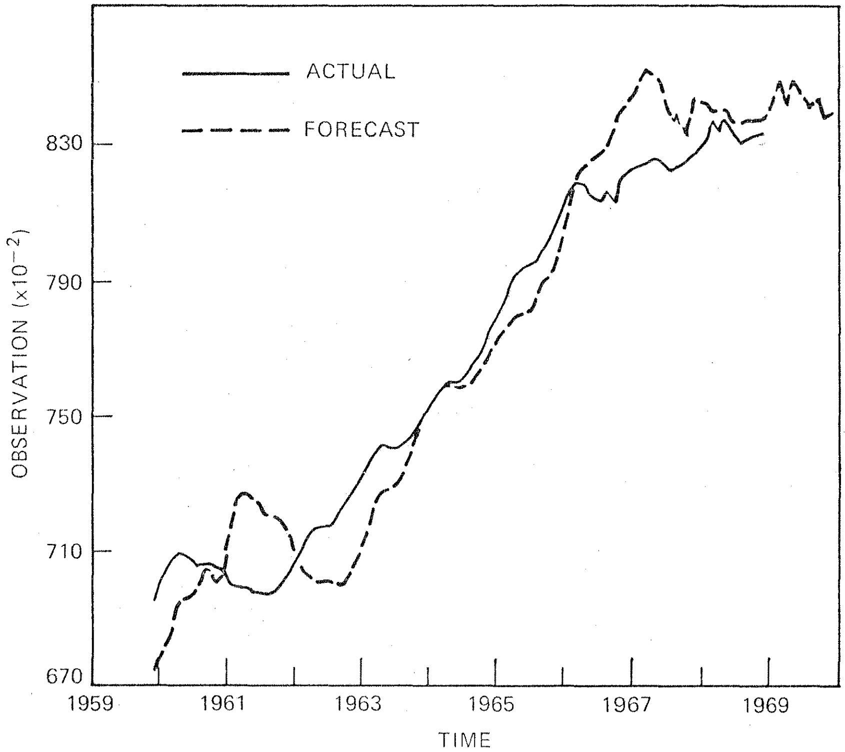
On the other hand, when forecast errors tend to alternate signs and be of approximately the same magnitude, the *SE* tends to zero while the SAE tends to a non-zero value. Thus, the absolute value of the tracking signal approaches zero, yielding a forecasting equation that gives a large weight to the smoothed history. The rate at which the smoothing constant adapts to changes in patterns of the forecast errors is a function of γ, the adaptive smoothing constant.

The choice of the best value of γ raises a question which is common to many forecasting schemes: should one choose a large value of γ to adapt to secular changes which occur in the series, or a lower value of γ which yields stable forecasts in the face of random fluctuations in the series being forecast? One way to resolve this dilemma is to choose the value of γ which minimizes the forecast error criteria evaluated for each of the local areas.

The use of adaptive exponential smoothing led to a sequence of local-area forecasting models. The first model developed assumed that the time series were made up of a trend and seasonal effect as in equation (1). The residual series, formed by subtracting the estimated quadratic trend from the actual series [see equation (3)], was forecast using the adaptive procedure. Unfortunately, this model suffered

FIGURE 5

FORECAST OF MAIN STATIONS FOR FLINT, MICHIGAN



from large and continuing forecast errors due to the inability of the quadratic trend to adapt to the dynamic shifts found in the local-area time series.

In an effort to have a model which was more responsive to shifts in the trend, the across-month exponential smoothing models of Winters[7] were extended. Winters' original models assumed that in a given series, $Z(t)$, the trend and seasonal effects are multiplicative, and he used simple exponential smoothing to smooth both the across-month component (trend), A_t, and the separate-month component (seasonal), I_t. As modeled for the local-area data, the across- and separate-month effects are assumed to be additive as in equation (1). The across-month effect is forecast using adaptive exponential smoothing:

$$A_t(Z) = \alpha(t)[Z(t) - I_{t-12}(Z)] + [1 - \alpha(t)]A_{t-1}(Z) . \qquad (11)$$

The separate-month effect is forecast using simple exponential smoothing by

$$I_t(Z) = \beta[Z(t) - A_t(Z)] + (1 - \beta)I_{t-12}(Z) . \qquad (12)$$

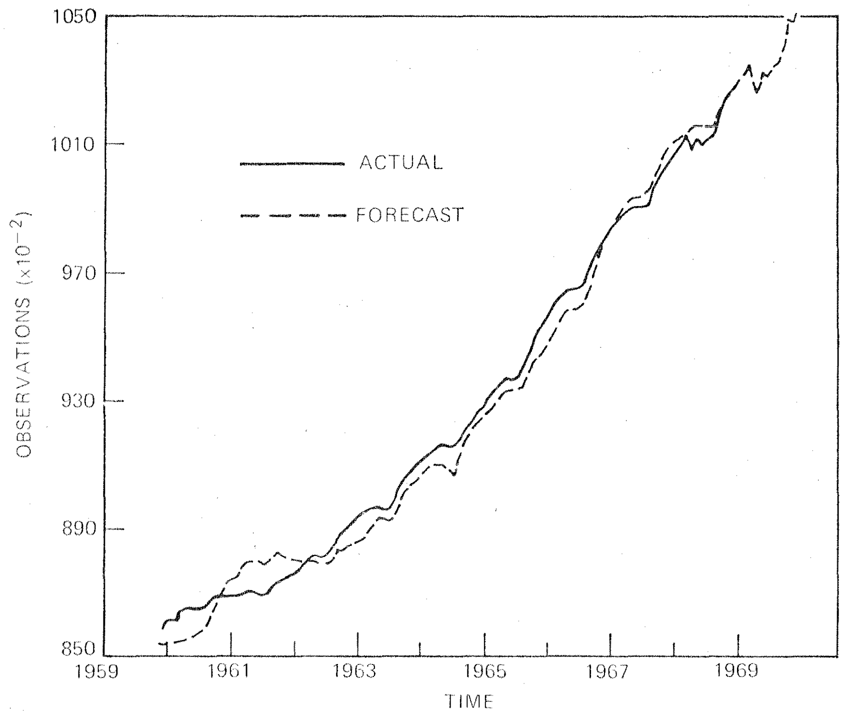
Note that the across-month effect is updated with each new observation. However, the separate-month factor for any particular month is updated only every twelve months as a new observation becomes available for that given month. Using this "two-way" smoothing procedure, a forecast is made by summing the seasonal and across-month effects for the appropriate lead time:

$$\hat{Z}(t + L) = \begin{cases} A_t(Z) + I_{t+L-12}(Z) , & \text{for } L \leq 12 , \\ A_t(Z) + I_{t+L-24}(Z) , & \text{for } 12 < L \leq 24 . \end{cases} \qquad (13)$$

[7] See [6].

FIGURE 6

FORECAST OF MAIN STATIONS FOR GRAND RAPIDS, MICHIGAN



It does not appear feasible, at this stage in the model development, to use adaptive exponential smoothing to forecast both the seasonal and across-month effects, although this appears to be an interesting matter for future research.

Therefore, since the seasonal fluctuations in the data appear to be more stable than the trend component, the adaptive procedure is used in the across-month effect. Note that simple exponential smoothing will adjust to changes in the seasonal level. However, this adjustment will not be as rapid as that given by the adaptive smoothing.

The forecasting procedure summarized by equation (13) is implemented by forecasting separately the connects and disconnects to yield a forecast of main station series. Shown in Table I are the parameter values which resulted in the lowest forecast errors in the main station series. Although not shown in this table, the connect and disconnect forecasts indicate that further improvements in forecast accuracy may be possible.

A third estimation procedure using adaptive exponential smoothing was developed which decomposed the connect or disconnect series, $y(t)$, into twelve monthly series, $y_j(t)$. The adaptive smoothing is used to forecast each of the monthly series separately,

$$\hat{y}_j(t+G) = S_t(y_j)\,, \quad G = \begin{cases} \dfrac{L}{12} & \text{for } L = 12, 24, 36, \ldots \\ \left[\dfrac{L}{12}\right] + 1 & \text{otherwise}\,. \end{cases} \tag{14}$$

Shown in Figures 3, 4, and 5 are the results from Flint, using adaptive exponential smoothing for the separate monthly series, each with a twelve-month lead time. The forecast main station series results from aggregating the forecast connects and disconnects. It can

TABLE 2

BEST FORECASTS FOR ADAPTIVE WITHIN–MONTH MODEL, EQUATION (14)

AREA	TYPE OF SERIES	LEAD TIME	GAMMA	FORECAST ERROR: AVERAGE ABSOLUTE	FORECAST ERROR: ROOT–MEAN–SQUARE
FLINT	CON.	12	0.50	182.6	242.4
	DIS.	12	0.50	144.9	195.9
	M.S.	12	0.50	1145.3	1454.3
	CON.	24	0.10	208.9	961.9
	DIS.	24	0.10	221.7	267.8
	M.S.	24	0.10	2030.0	2317.5
GRAND RAPIDS	CON.	12	0.40	130.9	174.3
	DIS.	12	0.40	108.9	136.2
	M.S.	12	0.40	498.3	567.1
	CON.	24	0.70	165.6	209.4
	DIS.	24	0.70	138.8	170.3
	M.S.	24	0.70	1018.1	1158.2
BATTLE CREEK	CON.	12	0.99	58.2	85.5
	DIS.	12	0.99	58.1	74.8
	M.S.	12	0.99	291.9	378.0
	CON.	24	0.99	76.1	99.9
	DIS.	24	0.99	64.7	91.0
	M.S.	24	0.99	754.0	864.9

NOTE:
THE VALUES PRESENTED ARE BEST IN THE SENSE THAT THEY MINIMIZE THE AVERAGE ABSOLUTE FORECAST ERROR. IN THE CASE WHERE THE AVERAGE ABSOLUTE FORECAST ERROR APPEARS INSENSITIVE TO MODEL PARAMETER CHANGES, THE VALUE PRESENTED IS THAT WHICH MINIMIZES THE ROOT–MEAN–SQUARE FORECAST ERROR. IN EACH CASE, N = 180 MONTHS AND B = 60 MONTHS. (CON., DIS., AND M.S. ARE ABBREVIATIONS FOR CONNECTS, DISCONNECTS, AND MAIN STATIONS RESPECTIVELY.)

be seen by visual inspection that the sudden changes in the series occurring in 1960–1962 and 1966–1969 are major sources in the forecast error for the connect, disconnect, and main station series forecasts. Although unreasonable forecast errors do not continue for long periods of time, there does not appear to be sufficient information in the series to forecast the sudden shifts which occur in the form of the series.

The main station forecast (with a twelve-month lead time) for Grand Rapids is shown in Figure 6. Note that the demand pattern exhibited by Grand Rapids is much more stable than that for Flint. As a result, the forecast errors for Grand Rapids are smaller than for Flint, even though the total number of main stations in service is larger by a factor of 1.5. Forecasting results for the three local areas, obtained by the adaptive within-month model, are shown in Table 2. In summary, the adaptive model works very well under stable conditions and reacts quickly to secular changes in the structure of the series. However, major shifts in the form of the series create large forecast errors during these periods of change.

3. Forecasting models using exogenous data

■ As stated above, although adaptive forecasting produced good results during periods of relative stability, there is not sufficient information in the series to forecast the major changes in the trend. On the other hand, the techniques of adaptive exponential smoothing produced good results in the face of the strong seasonal pattern shown in the connect-disconnect series. Consequently, it appeared reasonable to look for useful exogenous data to incorporate into the adaptive smoothing model. For exogenous data to be useful in forecasting applications, they should have the property that they "lead" the demand for residential telephone service in the local area being forecast.

The use of exogenous data in the forecast model introduces two related major problems. First, one has to decide which of all the conceptually useful exogenous variables are likely to be fruitful. Secondly, these data must be obtained with a reasonable expenditure of time and resources. For residential telephone demand in small and medium-sized metropolitan areas, one might judge that new household formation is closely related to new demand for residence main telephones. While it is true that demand for a second main station within the same household is rising, this demand is still very small compared to that generated by new households.

For one of the local areas studied (Flint), two exogenous series were obtained. These are (1) a monthly count of the total number of employees covered by unemployment compensation, and (2) a yearly series of the total number of households in the local area of Flint for 1960 through 1968.[8] In order to use the yearly exogenous series in the forecasting of a monthly series, the yearly series were broken down into monthly series by assuming that the growth was uniform throughout each year.

A number of models were developed to exploit this exogenous information, the most successful one being given in equation (15). Notice that the change in the exogenous series is used, and not the actual count. It is the change (gain or loss) which will effect the new demand for telephone service:

$$\hat{y}_j(t+G) = S_t(x, y_j) = \alpha(t)[y_j(t) + \beta\Delta x(t)] + [1-\alpha(t)]S_{t-1}(x, y_j)\,, \quad (15)$$

where

$$G = \begin{cases} \dfrac{L}{12} & \text{for } L = 12, 24, 36, \ldots \\ [L/12] + 1\,, & \text{otherwise}\,, \end{cases}$$

and

$$\Delta x(t) = x(t) - x(t-1)\,.$$

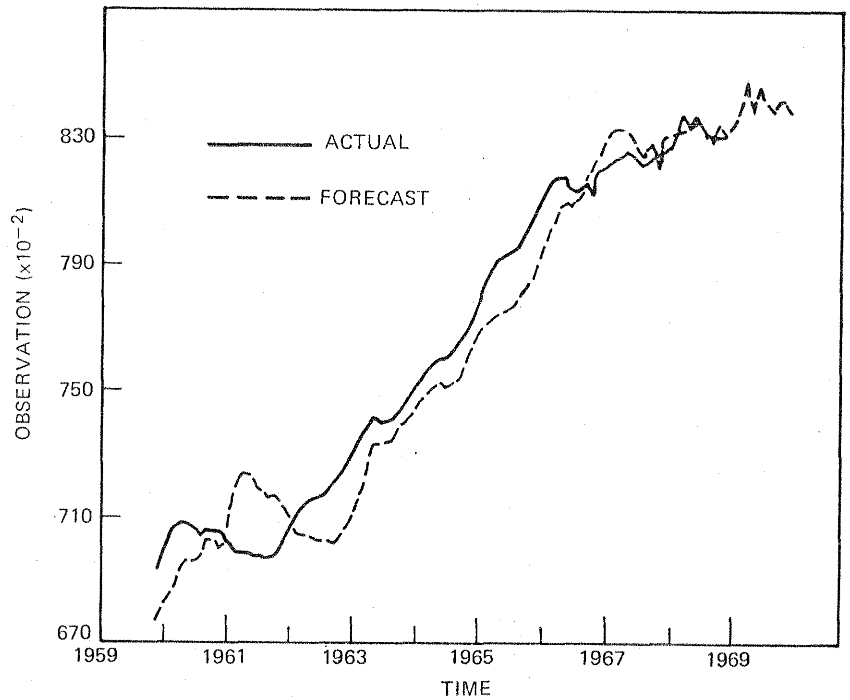
In this model, it is assumed that the exogenous data should also be weighted by the adaptively assigned $\alpha(t)$. Therefore, when changes in the form of the series do occur and forecast errors are of the same sign the new observation *and* the exogenous series will receive a large weight.

Among all the models studied, the results from using equation (15) to forecast the connect and disconnect series exhibit the largest reduction in main station forecast errors. Shown in Figure 7 are the best twelve-month main station forecasts for Flint using the exogenous series of households. The average absolute forecast error observed is 986.3, which is a reduction of approximately 15 percent over the best forecast generated without the exogenous data. It seems reasonable that if better exogenous data on new household formations

[8] Series (1) was obtained from the Michigan Employment Security Commission and series (2) from the Michigan Bell Telephone Company.

FIGURE 7

FORECAST OF MAIN STATIONS FOR FLINT, MICHIGAN USING EXOGENOUS DATA



were used, even further improvements in the forecast accuracy could be achieved.

4. Some conclusions and future plans

In a study such as this, which has ranged over a large number of different statistical techniques and over various sets of data, there are several useful conclusions that can be derived. Some of the more important of these are given below:

(1) The major problem of local-area forecasting is one of timing the construction of a new wire center. Consequently, it appears that analysis of historical time series should be made on the basis of groups of wire centers, where each group covers a meaningful geographical area. Gathering and analyzing such time series appears to be straightforward, and would be except for the fact that these apparently innocent time series of main stations can have treacherous characteristics. One of these characteristics stems from the fact that customers can be transferred from one wire center to another. Most often this happens when a new wire center is put into operation, but it can happen at other times also. This means that the historical time series for a specific wire center may refer to a variable geographical area, so that some of the apparent changes in demand simply reflect changes in the assignment of physical plant. Complete historical records of such changes are often not available.

(2) The business-residence split is necessary for statistical analysis. The characteristics of these two types of demand are sufficiently different that they should not be aggregated. As an example, residence gain seems to have a more regular trend than business gain. The latter is probably very much related to fluctuations in both general and local business conditions.

(3) The characteristics of the connect and disconnect series are sufficiently different that they also need to be analyzed separately. As an extreme example of this, these two series may have substantially different patterns in resort areas. A result of these differences is that a combined series of total main stations not only has larger variability but is also harder to analyze and comprehend.

(4) Some short-term local swings are almost impossible to predict by present methods. Local swings are frequently brought about by strikes, welfare policy, zoning changes, and deposit practices. Swings such as these may not be predictable by any statistical method whatsoever.

(5) While the analysis of the data and the improvements in the models have increased the forecasting accuracy, the forecasts need to be better yet. If empirical confidence limits on the forecasts are developed[9] and translated into intervals of time, then it becomes clear that the difference between the earliest and latest times at which new plant will be required is often too large to be of any practical value.

(6) The 15-percent gain in accuracy achieved in the Flint forecasts by the introduction of the exogenous data is an important indicator of further gains that can be made by the use of relevant exogenous data. After all, this gain was achieved by the introduction of only nine annual observations on new household formations. In addition, the city of Flint was selected because of the apparent difficulty in forecasting there. The statistical characteristics of Grand Rapids are much more regular. This result tends to verify our original feelings that analysis of the historical time series of main stations is an important step towards better local area forecasts, but that a more direct study of the factors which generate the local demand for telephone service is needed, i.e., a search for closely related exogenous data. Such a search will not be easy because the potential demand factors are many and complex. For example, the characteristics of demand are not likely to be the same in urban and in suburban areas. In surburbia, the demand is perhaps related to the number of families moving into new homes and apartments, while in the more urban areas the demand for telephones may be related to shifting land use. Other characteristics of demand will undoubtedly be important, and all of them are probably dynamic in that they are closely related to population mobility characteristics.

In summary, then, five main directions appear promising for future work:

(1) Obtain and analyze more accurate historical information by wire centers, including both exogenous and main station time series.

(2) Survey customer demand by local area, including the manner in which it relates to population mobility and dynamic customer characteristics.

(3) Develop models specifically to forecast the time at which plant capacity is exceeded.

---

[9] See Williams and Goodman [5].

(4) Examine alternative error criteria for parameter estimation which more fully represent the costs associated with forecast errors.

(5) Forecast other facets of plant utilization, such as equipment usage, rather than simply equipment demand.

References

1. Box, G. E. P. and Jenkins, G. M. *Time Series Analysis*, San Francisco: Holden-Day, Inc., 1970.
2. Chaddha, R. L. and Chitgopekar, S. S. "A 'Generalization' of the Logistic Curves and Long-Range Forecasts (1966–1991) of Residence Telephones," *The Bell Journal of Economics and Management Science*, this issue.
3. Chambers, J. M. "Regression Updating," *Journal of the American Statistical Association*, to appear (December 1971).
4. Trigg, D. W., and Leach, A. G. "Exponential Smoothing With an Adaptive Response Rate," *Operational Research Quarterly*, Vol. 18, No. 1 (1967), pp. 53–59.
5. Williams, W. H. and Goodman, M. L. "A Simple Method for the Construction of Empirical Confidence Intervals for Economic Forecasts," *Journal of the American Statistical Association*, to appear (December 1971).
6. Winters, P. R. "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, Vol. 6, No. 3 (1960), pp. 324–42.