THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 61, No. 1, January 1982
Printed in U.S.A.

A Robust Sequential Projection Algorithm for Traffic Load Forecasting

By J. P. MORELAND

(Manuscript received December 31, 1980)

Forecasts of busy season trunk group traffic loads are required for planning the Bell System's message network. Forecasting algorithms currently in use obtain estimates of future loads by multiplying the most recent measurement of busy season load by an aggregate growth factor. Because of statistical errors in measured loads and differences between individual trunk group and aggregate growth factors, the resulting forecasts can have large statistical errors. In this paper we extend earlier work to develop a new algorithm, called the sequential protection algorithm (SPA), based on a linear two-state Kalman filter, together with logic for detecting and responding to unusually large measurement errors or changes in trend. In typical applications of Kalman filtering, the statistics of system noises, measurement errors, and initial conditions are known and the filter parameters (Kalman gains) are selected accordingly. For our application, however, these statistics cannot be determined without error. Consequently, we develop a method for selecting robust filter parameters which provide improved performance, independent of system noises, measurement errors, and initial conditions. In particular, under the assumption of linear growth for 5-year intervals, the average rms 1-year forecast error of SPA is about 10 percent less than that of the existing algorithms. Field test results confirm the theoretical results presented here. Accordingly, specifications have been written for inclusion of SPA in the Bell System's standard trunk forecasting systems.

I. INTRODUCTION AND SUMMARY

Forecasts of busy season (yearly peak) trunk group traffic-loads are required for the planning of the Bell System's message network. These forecasts are used to design traffic networks which minimize the cost of the trunks required to satisfy anticipated customer demands.

The standard load forecasting algorithms currently in use in the Bell System obtain estimates of future loads by multiplying the most recent measurement of busy season load by an aggregate growth factor; for example, the average of the growth factors obtained by trending the total office loads at each end of the trunk group. Descriptions and comparisons of the various algorithms currently in use are given in Ref. 1.

As explained in Ref. 1, these algorithms have two significant sources of error: (*i*) Because of the finite amount of data upon which measurements are based, measured loads can have large statistical errors; standard deviations fall in the range of about 5 to 40 percent depending upon load size and type of measurement system.[2] (*ii*) Individual trunk group growth factors can differ from the aggregate growth factor; standard deviations of 6 percent have been observed. These forecast errors are significant since they lead to an increase in the reserve trunk capacity required to satisfy customer demands.[3]

To reduce forecast error and, hence, reserve trunking capacity, a new algorithm, called the sequential projection algorithm (SPA), has been developed to forecast busy season traffic loads within the Bell System. The SPA is based on a linear two-state Kalman filter model, whose use in traffic forecasting was studied first by David and Pack,[1] together with logic for detecting and responding to outlier measurements, i.e., unusually large measurement errors or changes in trend.

As discussed in Ref. 1, David and Pack tested several Kalman filter models—some with as many as eight state variables and four data variables. In summary, for the planning interval of interest (1 to 5 years ahead) none performed consistently or significantly better than the relatively simple two-state (traffic load and incremental growth), one-data variable (measured load) model.

The two-state Kalman filter establishes a linear trend for individual traffic loads as follows: As illustrated in Fig. 1, the level, or smoothed base load, is a weighted average of the most recent measurement, or base load, and the previous 1-year forecast. Similarly, the smoothed growth increment is a weighted average of the measured and previously forecasted increments. (The measured increment is, by definition, the measured load minus the previous year's smoothed base load.)

The performance of the Kalman filter, as measured by mean square forecast error, depends upon the filter parameters, i.e., the gains α_n and β_n in Fig. 1, the standard deviation of load measurement and growth estimation errors, and upon the assumed evolution of the true load.

Since SPA will be used under a variety of possible operating conditions, the filter parameters could be tuned to provide optimal performance, i.e., minimum mean square forecast error, for each application.

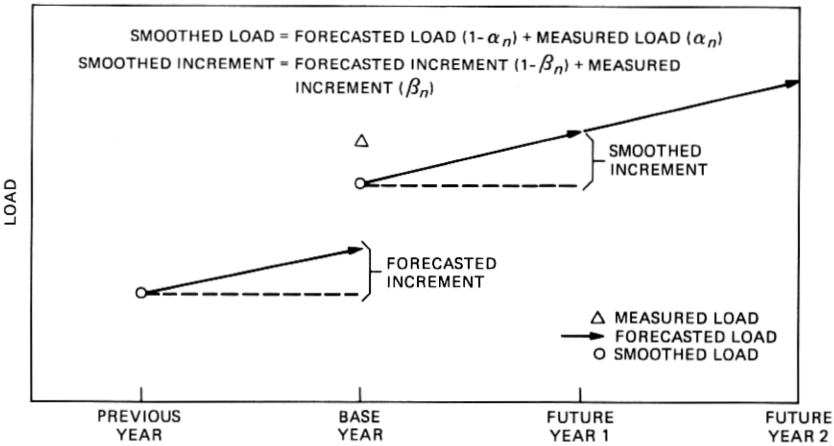

Fig. 1—Basic operation of SPA.

For example, in the Bell System we employ two types of load measurement systems; Bell Operating Companies obtain load estimates from a direct measurement of trunk group usage, while Long Lines derives estimates from point-to-point, e.g., end-office to end-office data provided by the Centralized Message Data System (CMDS).[2] For trunk group data, the standard deviation of measurement error is in the range of about 5 to 10 percent, depending on load size. For point-to-point data, the range is about 10 to 40 percent. (See Fig. 2.) Accordingly, different parameters could be used for different measurement systems and for different load ranges.

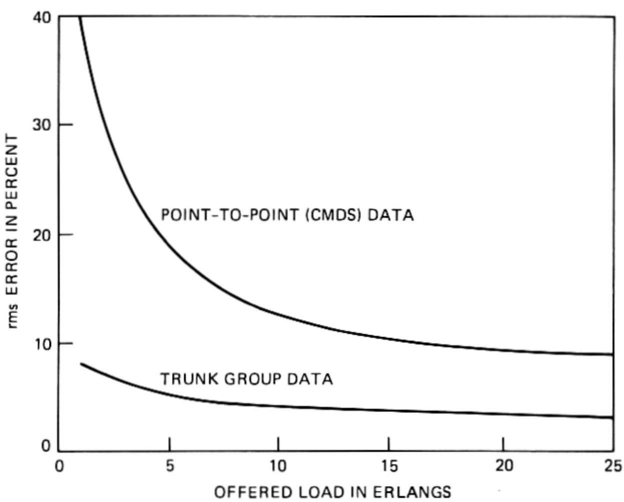

Fig. 2—Sampling error vs. offered load.

Indeed, the initial development of SPA[1] was intended for use with trunk group data only, and the objective of the current study was to extend the original work to applications using point-to-point data. As suggested above, one possible solution would be to develop multiple versions of SPA.

Instead, in this paper, we develop a single, robust SPA whose parameters are selected to provide improved performance over the entire range of operating conditions, including the use of either trunk group or point-to-point data.

The use of robust parameters is important for two reasons: (*i*) A single SPA for all applications should be simpler to implement and maintain than multiple versions. (*ii*) The actual values of the statistical parameters for each application cannot be determined without error, and our results show that erroneously assumed values can lead to a performance substantially worse than that of conventional projection methods. Of course, as with any robust technique, we pay a premium by receiving less than theoretically optimal performance for protection against the possibility of a performance worse than conventional methods.

Section II of this paper provides a qualitative overview of the functions performed by SPA that include procedures for detecting and

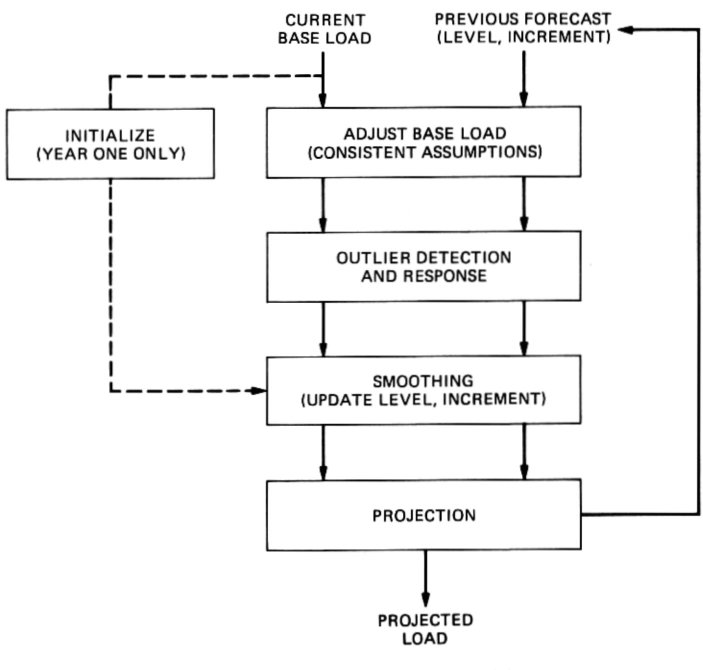

Fig. 3—Sequential projection algorithm.

responding to outliers, as well as the smoothing and projection functions of the Kalman filter. Section III defines the mathematical model and procedures for selecting robust filter parameters; Section IV gives numerical results; Section V develops the outlier detection thresholds; and Section VI gives the conclusions.

II. OVERVIEW OF SPA

As shown in Fig. 3, SPA is composed of five major components: algorithm initialization, base-load adjustments, outlier detection, smoothing, and projection. In the following paragraphs, we provide a qualitative description of the operations performed by each of these components.

2.1 Initialization

As indicated in Fig. 4, the smoothed base load in the first year of operation, and in certain other cases discussed in Section 2.3, is equal to the measured base load. The smoothed growth increment is calculated by multiplying this base load by a growth factor obtained, for example, by trending the total office loads at each end of the trunk group.[1]

2.2 Base load adjustments

In the second and subsequent years of operation, SPA updates both the level and incremental growth by comparing the most recent base load with the previous 1-year forecast of that same load. Since the base and forecasted loads must correspond to the same traffic routings,

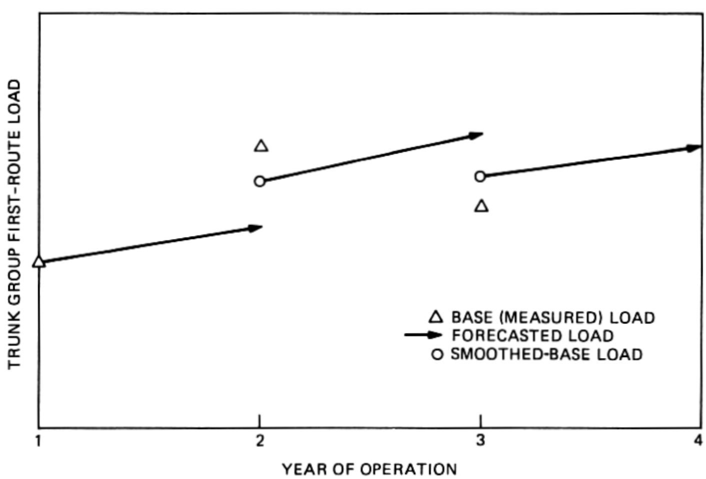

Fig. 4—Initialization of SPA.

differences between the previous and current routings are accounted for by adding an adjustment term to the base load so that the adjusted base load agrees with previous routing assumptions. After performing the outlier detection and smoothing functions described below, the routing adjustment term is subtracted from the smoothed base load so that it agrees with the current routing assumptions.

Furthermore, the previously forecasted load is adjusted to remove the impact of deterministic events, such as a proposed tariff change, that were predicted but did not occur.

2.3 Outlier detection

Under the assumption that the observed forecast error (i.e., the difference between the forecasted and measured loads) has a Gaussian distribution with zero mean, the linear Kalman filter upon which SPA is based provides the minimum attainable mean squared forecast error.[4] In practice, however, the normal statistical errors (because of the finite measurement interval, day-to-day load variation, random variations in CMDS point-to-point sample size,[2] and growth errors) are occasionally contaminated by wiring errors, recording errors, or unexpected changes in the trend of the true load. In such cases, when the observed forecast error deviates from a Gaussian distribution, the linear Kalman filter model can have a mean squared forecast error which is substantially greater than the minimum.[4]

Our approach to this problem, which is based in part on the nonlinear

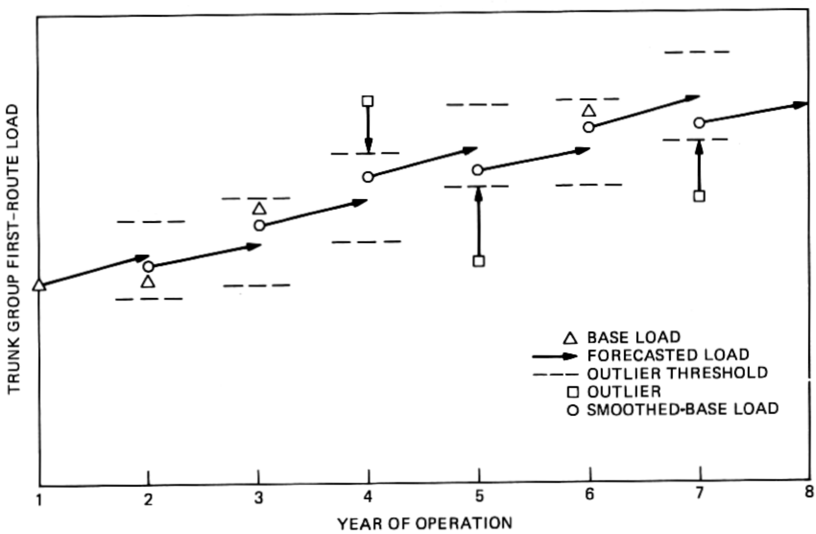

Fig. 5—Response of outliers (sign not repeated).

Kalman filter model described in Ref. 4, is the following: If the difference between the possibly adjusted base and forecasted loads exceeds present thresholds, the base load is declared to be an outlier. In response, SPA will adjust the base load or restart at the measured level, depending only upon whether an outlier of the same sign occurred in the previous year.

If an outlier of the same sign did not occur in the previous year, SPA replaces the base load by the nearest threshold value as indicated by the vertical arrows in Fig. 5. Qualitatively, the underlying assumption here is that in most cases such an outlier signals an invalid or atypical measurement caused by, for example, a recording error and not a change in trend. Formally, this modification of the Kalman filter is equivalent to that proposed by Masreliez and Martin in Ref. 4.

Alternatively, as indicated in Fig. 6, if an outlier of the same sign occurs in two consecutive years, SPA restarts at the measured level. The assumption here is that in most cases two consecutive outliers of the same sign signal a change in trend. In theory, additional improvement could be obtained by restarting at the previous outlier and then smoothing with the current measurement. However, such a procedure would be more complicated to implement, and our studies show that it would have negligible impact on performance.

As indicated by Huber's studies,[5] and as supported by our field test results,[6] adequate protection against outliers is obtained when the thresholds are set anywhere in the range of about one or two times the

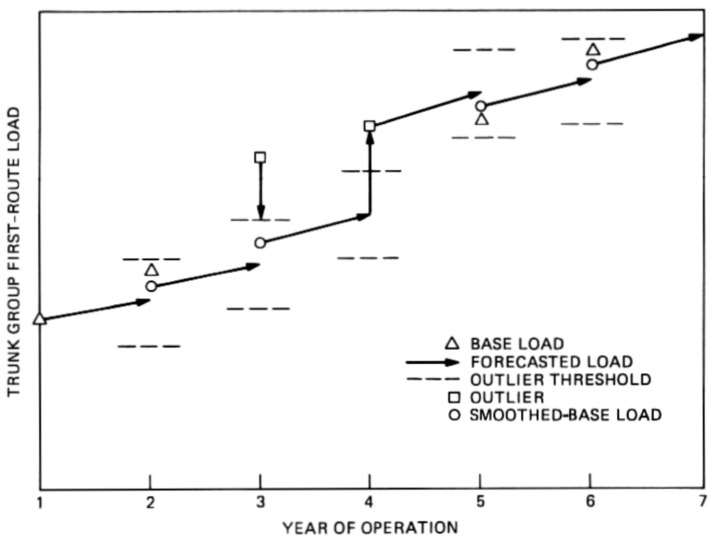

Fig. 6—Response to outliers (sign repeated).

rms observed forecast error. This result is important: we show that it allows us to use thresholds which are independent of the number of data points processed and the type of measurement system. Numerical values for these thresholds are provided in Section V.

2.4 Smoothing

The possibly adjusted base and forecasted loads are combined to produce a smoothed base load and a smoothed growth increment. As indicated in Fig. 1, the smoothed base load is a weighted average of the base and 1-year forecasted loads. Similarly, the smoothed growth increment is a weighted average of the measured and forecasted growth increments. Procedures for selecting the appropriate gains α_n and β_n are described in Section IV.

2.5 Projection

As indicated in Fig. 1, the smoothed base load and growth increment are combined to establish a linear projection. That is, the projected load for the kth future year is obtained by adding k smoothed growth increments to the smoothed base load. In practice, the resulting trunk group load forecasts can be adjusted to include user supplied estimates of the impact of deterministic events (caused by, for example, proposed tariff changes) or to agree with other aggregate load forecasts.

III. MATHEMATICAL MODEL

3.1 General

Although SPA is based on a linear two-state, i.e., trunk group load and incremental growth, Kalman filter, we considered more complex models with additional state variables, e.g., aggregations of other trunk group loads and growth factors. Therefore, for completeness, we first summarize the equations which define the general discrete-time linear Kalman filter. For a more complete discussion, see Ref. 7.

The true time-behavior of the state variables is assumed to be defined by the linear transition equation

$$\underline{X}_{n+1} = \phi \underline{X}_n + \underline{W}_n + \underline{U}_n, \tag{1}$$

where $\underline{X}_n$ is an s-vector of true state variables in period (year) n, ϕ is an $s \times s$ transition matrix which may depend on n, $\underline{W}_n$ is an s-vector of zero-mean random modelling errors, and $\underline{U}_n$ is an s-vector of deterministic changes in state.

The one-period-ahead projection formula is

$$\underline{X}_{n+1,n} = \phi \underline{X}_{n,n} + \underline{U}_n, \tag{2}$$

where $\underline{X}_{n+k,n}$ denotes an estimate of $\underline{X}_{n+k}$ based on data through period n. The "smoothed" estimate $\underline{X}_{n,n}$ is calculated by

$$\underline{X}_{n,n} = \underline{X}_{n,n-1} + K_n(\underline{y}_n - H\underline{X}_{n,n-1}), \tag{3}$$

where K_n is a $d \times s$ "Kalman gain" matrix and $\underline{y}_n$ is a d-vector of measurements in period n related to $\underline{X}_n$ by

$$\underline{y}_n = H\underline{X}_n + \underline{V}_n, \tag{4}$$

where H is a $d \times s$ matrix and $\underline{V}_n$ is a d-vector of zero-mean measurement errors.

The optimal gain matrix K_n is determined recursively by the equations

$$K_n = P_n H^T (HP_n H^T + R)^{-1} \tag{5}$$

$$S_n = (I - K_n H)P_n, \tag{6}$$

and

$$P_{n+1} = \phi S_n \phi^T + Q, \tag{7}$$

where R is the covariance matrix of measurement errors, i.e., $R = E(\underline{V}_n \underline{V}_n^T)$, Q is the covariance matrix of modelling errors, and S_0 is an estimate of the covariance matrix of the initial state estimate $\underline{X}_{0,0}$. Furthermore, it is assumed that $\underline{V}_i$ and $\underline{W}_j$ have zero mean and are pairwise uncorrelated among themselves and with each other for all i, j.

As discussed in Section I, we will be concerned with filter performance for nonoptimal gains. In this case, the covariance matrix P_{n+1} of the projection error $(\underline{X}_{n+1,n} - \underline{X}_{n+1})$ is related to the covariance matrix S_n of $\underline{X}_{n,n}$ by eq. (7), but S_n is related to P_n by

$$S_n = (I - K_n H)P_n(I - K_n H)^T + K_n R K_n^T, \tag{8}$$

which reduces to eq. (6) only when K_n is given by eq. (5).

3.2 *Two-state model*

As noted in Section I, we considered models with as many as $s = 8$ state variables and $d = 4$ measurement variables. However, our studies[1] showed that none performed consistently or significantly better than the simple two-state, one-data variable model defined by the equations

$$\underline{X}_{n+1} = \begin{bmatrix} x_{n+1} \\ \dot{x}_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ \dot{x}_n \end{bmatrix} + \begin{bmatrix} w_n \\ \dot{w}_n \end{bmatrix} \tag{9}$$

and

$$y_n = x_n + v_n, \tag{10}$$

where x_n and $\dot{x}_n$ are, respectively, the true load and true incremental growth in year n, and y_n is the measured load in year n. These equations correspond to eqs. (1) and (4) with

$$\phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \tag{11}$$

and

$$H = [1, 0]. \tag{12}$$

The more complex models were rejected since the additional variables that we considered, e.g., aggregations of other trunk group loads and growth factors, are usually not highly correlated with trunk group load. Moreover, when the correlation is high the improvement in performance is minimal and, more importantly, when high correlation is incorrectly assumed, the penalties outweigh the expected benefits.[1]

To complete the description of our two-state model, note that the smoothing eq. (3) becomes

$$\underline{X}_{n,n} = \begin{bmatrix} x_{n,n} \\ \dot{x}_{n,n} \end{bmatrix} = \begin{bmatrix} x_{n,n-1} + \alpha_n(y_n - x_{n,n-1}) \\ \dot{x}_{n,n-1} + \beta_n(y_n - x_{n,n-1}) \end{bmatrix}, \tag{13}$$

where α_n and β_n are the Kalman gains in year n. The 1-year-ahead projection formula is

$$\underline{X}_{n+1,n} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{n,n} \\ \dot{x}_{n,n} \end{bmatrix}. \tag{14}$$

Note: Our analysis of the two-state model will ignore the possibility of deterministic changes in state; that is, we assume $\underline{U}_n = 0$. Also, by combining eqs. (13) and (14) it follows that eq. (13) may be written in the form

$$\begin{bmatrix} x_{n,n} \\ \dot{x}_{n,n} \end{bmatrix} = \begin{bmatrix} (1 - \alpha_n)x_{n,n-1} + \alpha_n y_n \\ (1 - \beta_n)\dot{x}_{n,n-1} + \beta_n(y_n - x_{n-1,n-1}) \end{bmatrix}, \tag{15}$$

which emphasizes that $x_{n,n}$ is a weighted average of the forecasted and measured levels and, similarly, $\dot{x}_{n,n}$ is a weighted average of the forecasted and measured increments.

Finally, note that $R = \sigma^2$, where σ is the standard deviation of load measurement error; the dependence of σ on load size and type of measurement system is discussed in Section 4.2.2. In the following section, we define the SPA initialization procedure and in Section 3.4 we describe our procedure for selecting values for the filter gains α_n and β_n and the associated assumptions about Q.

3.3 Initialization

As explained in Section 2.1, the smoothed-base load and growth increment in the first year of operation are given by

$$x_{0,0} = y_0 \tag{16}$$

and

$$\dot{x}_{0,0} = \hat{g} y_0, \tag{17}$$

where $\hat{g}$ is an aggregate growth factor. For example, if $\hat{g} = 0.1$, the increment is 10 percent of the base load.

The normalized covariance matrix $S_0/\sigma^2 = (s_{ij}/\sigma^2)$ of the initial state estimate is given by

$$\frac{s_{11}(0)}{\sigma^2} = \frac{E(x_0 - y_0)^2}{\sigma^2} = 1, \tag{18}$$

$$\begin{aligned} \frac{s_{22}(0)}{\sigma^2} &= \frac{E(gx_0 - \hat{g}y_0)^2}{\sigma^2} \\ &= \frac{E[(g - \hat{g})x_0 + \hat{g}(x_0 - y_0)]^2}{\sigma^2} \\ &= \frac{\sigma_g^2 x_0^2}{\sigma^2} + \hat{g}^2, \end{aligned} \tag{19}$$

where g is the true growth factor and σ_g is the standard deviation of the estimate $\hat{g}$, and

$$\begin{aligned} \frac{s_{12}(0)}{\sigma^2} &= \frac{E[(x_0 - y_0)(gx_0 - \hat{g}y_0)]}{\sigma^2} \\ &= \frac{E\{(x_0 - y_0)[(g - \hat{g})x_0 + \hat{g}(x_0 - y_0)]\}}{\sigma^2} \\ &= \hat{g}. \end{aligned} \tag{20}$$

3.4 Filter parameters

As discussed in Section I, our objective is to select robust values for the gains α_n and β_n; that is, values which will provide improved performance over the range of operating conditions. Our approach to this problem starts with the following idealistic assumption, which will be relaxed in a later section:

We first assume that the true load displays constant incremental growth. That is, we first assume that Q, the covariance matrix of modelling errors, is identically zero. Under this assumption, it follows from eqs. (5) to (7) that the optimal gains depend only on the ratio S_0/σ^2. In turn, for a given value of $\hat{g}$, it follows from eqs. (18) to (20) that S_0/σ^2 is determined by the ratio

$$G^2 = \frac{\sigma_g^2}{(\sigma/x_0)^2}. \tag{21}$$

In Section 4.2, we display the optimal gains and corresponding performance (as measured by mean square forecast error) for various known values of G. More importantly, we show that an erroneously assumed value for G can lead to a performance worse than that of the conventional projection method; equivalently, since the conventional method is used to initialize SPA, the mean square forecast error can increase with the number of data points processed.

Next, in Section 4.2.3, we show that the gains corresponding to one particular value of G provide a performance that is nearly independent of the actual value of G. We call these gains the "robust gains for linear growth."

Finally, in Section 4.3, we relax the assumption of linear growth and show that certain constant gains, derived from the "robust gains for linear growth," provide improved performance in the presence of system noise, as well as the ideal case where the trend remains constant.

IV. NUMERICAL RESULTS

4.1 Examples

To help explain our results, we first consider several examples which show how the gains and the mean-square forecast error depend upon the ratio G. In these examples, we assume that the true load has constant incremental growth, and we assume two-point distributions for measurement and growth estimation errors. That is, the measured load is either high or low by equal amounts and with equal probability, and the initial estimate of incremental growth is either high or low by equal amounts and with equal probability. Moreover, in each example we assume that the aggregate growth factor $\hat{g}$ is zero.

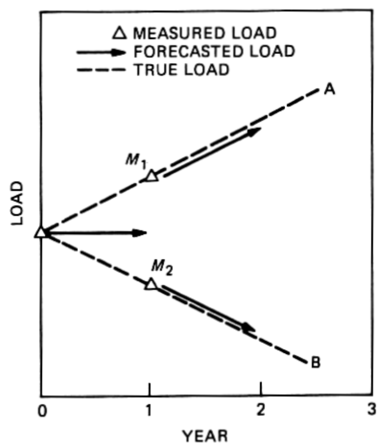

Fig. 7—Optimal filter: no measurement error ($G = \infty$).

4.1.1 No measurement error (G = ∞)

In the example illustrated in Fig. 7, we assume no measurement error, i.e., $\sigma^2 = 0$, but an error of plus or minus one unit in the initial increment so that $G = \infty$. Thus, with this information, the true load lies along the dashed line marked A with probability 0.5 or along the dashed line B with probability 0.5. Accordingly, the initial mean square forecast error is $0.5(1)^2 + 0.5(-1)^2 = 1$.

As shown in Fig. 7, the possible measurements in year 1 are M_1 or M_2 with equal probability. However, since there is no measurement error and since two points determine a straight line, it is clear that in either case the forecast error can be reduced to zero after processing the second measurement; examination of eq. (13) shows that the appropriate gains are $\alpha_1 = \beta_1 = 1$.

4.1.2 No growth error (G = 0)

At the other extreme, suppose there is no error in the initial growth estimate, i.e., $\sigma_g^2 = 0$, but an error of plus or minus one unit in the measured load so that $G = 0$. As shown in Fig. 8, the trend line is either A or B with equal probability; hence, the initial mean square forecast error is 1. The possible measurements in year 1 are M_1 with probability 0.25 [since A and a high measurement occur with probability (0.5)(0.5)], M_2 with probability 0.5, and M_3 with probability 0.25.

Since M_1 and M_3 correspond uniquely to A and B, respectively, the forecast error can be reduced to zero after processing either of these measurements; eq. (13) and Fig. 8 show that the appropriate gains are $\alpha_1 = 0.5$ and $\beta_1 = 0$. However, if M_2 occurs, no new information is obtained. Consequently, after processing the second data point, the

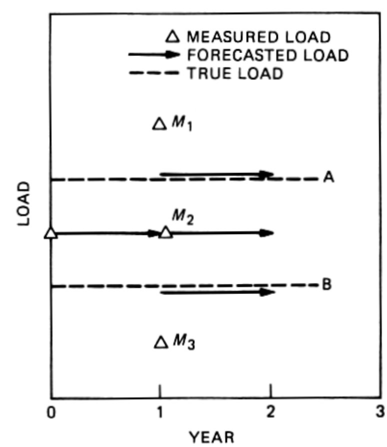

Fig. 8—Optimal filter: no growth error ($G = 0$).

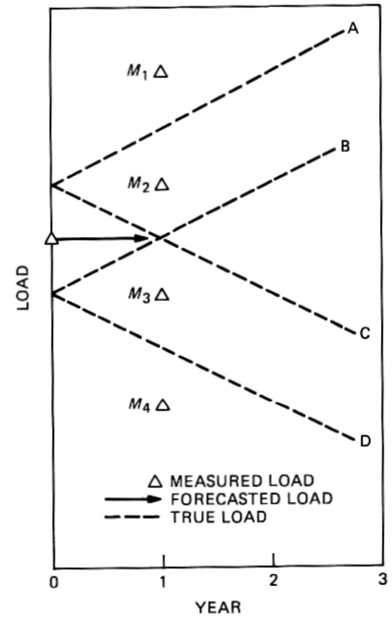

Fig. 9—Optimal filter: equal measurement and growth errors ($G = 1$).

mean square forecast error is $0.5(0)^2 + 0.5(1)^2 = 0.5$. Thus, the rms forecast error is reduced by a factor of $1/\sqrt{2}$.

4.1.3 Equal measurement and growth errors (G = 1)

For values of G between the above extremes, i.e., $0 < G < \infty$, the reductions in forecast error might be expected to fall between those corresponding to the extreme values of G. But, this is not generally true.

For example, consider the case illustrated in Fig. 9 in which $G = 1.0$. We assume uncorrelated errors of plus or minus one unit in the measured load and in the initial increment. Thus, as shown in Fig. 9, the possible trend lines are A, B, C, or D, each with equal probability. Consequently, the initial mean square forecast error is $0.25(2)^2 + 0.5(0)^2 + 0.25(-2)^2 = 2$.

In year 1 the possible measurements are M_1 with probability $[0.25(0.5) = 0.125]$, M_2 with probability 0.375 (since M_2 may correspond to A, B, or C), M_3 with probability 0.375, and M_4 with probability 0.125.

If M_1 occurs, the trend line must be A and the forecast error could be reduced to zero by setting $\alpha_1 = 2/3$ and $\beta_1 = 1/3$. Similarly, if M_4 occurs the trend line must be D; again $\alpha_1 = 2/3$ and $\beta_1 = 1/3$ are appropriate.

If M_2 occurs, the trend line is either A, B, or C with equal probability. Since B is halfway between A and C in year 2, it follows that the mean square forecast error is minimized by forecasting the level of B in year 2; this result is obtained with $\alpha_1 = 2/3$ and $\beta_1 = 1/3$. Similarly, if M_3 occurs, the forecast with these gains is on C in year 2, and the forecast error is minimized.

Thus, the optimal gains are $\alpha_1 = 2/3$, $\beta_1 = 1/3$ and, after processing the second data point, the mean square forecast error is $0.5(0)^2 + 0.25(2)^2 + 0.25(-2)^2 = 2$, which is identical to the initial value. Thus, for this example there is no reduction in forecast error after processing the second data point.

4.1.4 G unknown

The above examples assumed that G was known exactly. Suppose now, however, that the gains are chosen under the assumption that $G = \infty$, i.e., $\alpha_1 = \beta_1 = 1$, but in fact $G = 0$. In this case, which is illustrated in Fig. 10, the mean square forecast error in year 2 is $0.25(3)^2 + 0.25(1)^2 + 0.25(-1)^2 + 0.25(-3)^2 = 5$, which is five times the initial value.

This result, although exaggerated, is important because it shows that SPA could actually perform worse than the conventional forecasting procedure. Indeed, this fact, combined with the practical impossibility of estimating G without error, is the major consideration in our decision to use the robust version of SPA described in Sections 4.2.3 and 4.3.2.

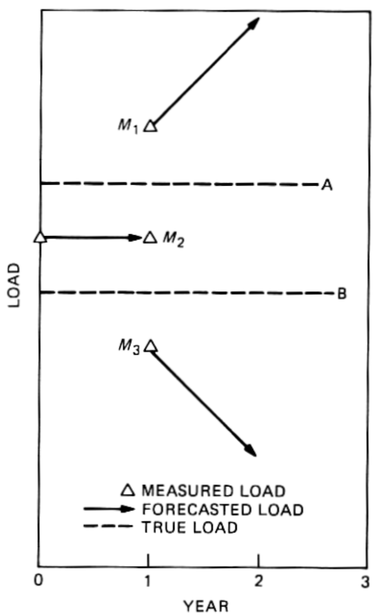

Fig. 10—Nonoptimal filter: assumed $G = \infty$, actual $G = 0$.

4.2 Linear growth

4.2.1 G known

The results of this section apply when the true load has constant incremental growth and when the ratio G, eq. (21), is known exactly.

Figure 11 displays the normalized rms 1-year forecast error as a function of the number of data points processed after initialization. The normalization is with respect to the rms forecast error of the conventional projection method. Since this method is used to initialize SPA, the initial normalized rms error equals 1.0. Each curve corresponds to a different value of G and, accordingly, to a different set of gains α_n and β_n. For example, the gains corresponding to three different values of G are shown in Fig. 12.

The results shown in Fig. 11 assume that the true load displays constant incremental growth; under this assumption, the forecast error approaches zero as n increases. In practice, however, unexpected changes in trend will occur and, consequently, as described in Section 2.3, SPA will occasionally be reinitialized. Accordingly, average forecast error over the interval between reinitializations is a more meaningful figure of merit. In the following paragraphs, we assume an average interval of five years. We emphasize, however, that this assumption will underestimate the benefits of SPA under more stable conditions and overestimate the benefits under less stable conditions.

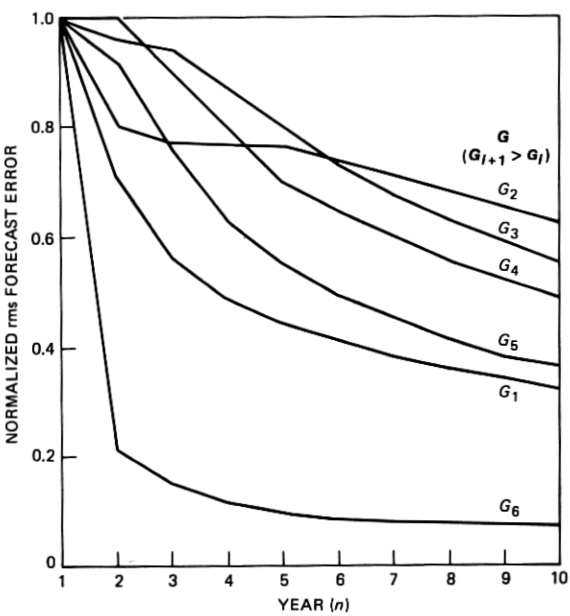

Fig. 11—Optimal filter: forecast error vs. year.

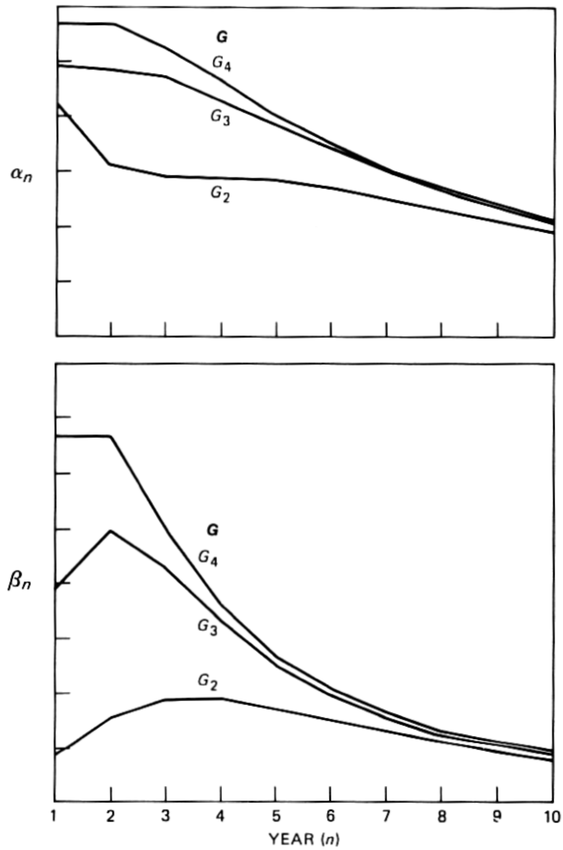

Fig. 12—Optimal gains vs. year.

Thus, the curve labeled optimal in Fig. 13 displays the normalized 5-year average forecast error—the average of the first five values in Fig. 11—as a function of the parameter G. If, for example, $G = G_2$ and we choose the gains accordingly, the average rms error would be about 20 percent less than that for the conventional forecasting method. Alternatively, if $G = G_4$, we would choose a different set of gains and the average reduction would be approximately 15 percent. Note as suggested by the examples discussed above, that the reduction in forecast error is a convex function of G.

4.2.2 G unknown

Suppose now that the actual value of G differs from the assumed value. For example, suppose we use the gains corresponding to $G = G_2$, but in fact $G = G_4$. In this case, the curve labeled $G = G_2$ in Fig. 13 shows that the 5-year average rms error would be about 20 percent

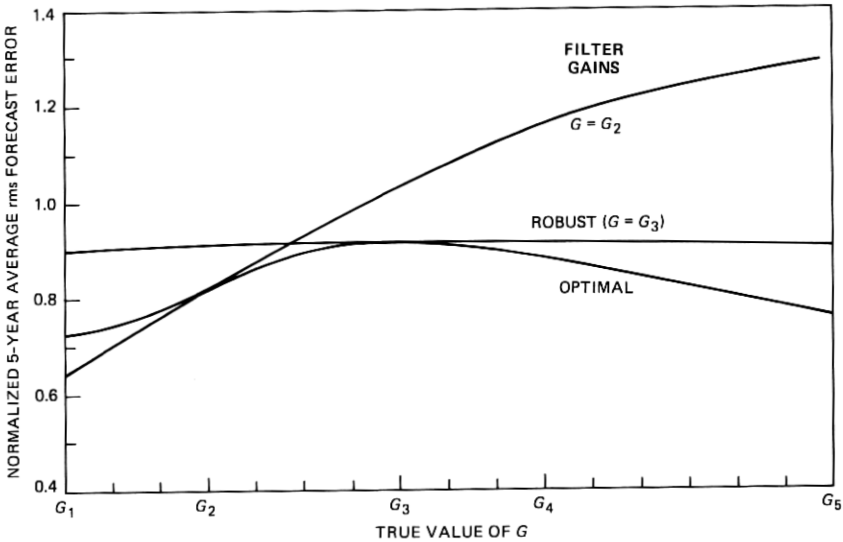

Fig. 13—Average rms forecast error for several filters.

larger than its initial value, in agreement with the example discussed in Section 4.1.4.

The results of Fig. 13 can also be interpreted as follows: Suppose that $G = G_4$ is appropriate when trunk group loads are derived from trunk group data. Then $G = G_2 = G_4/3$ would be appropriate when loads are derived from point-to-point (CMDS) data since, from Fig. 2, the rms measurement error for CMDS data is approximately three times that for trunk group data. Figure 13 then shows that an SPA tuned for point-to-point data would perform poorly with trunk group data.

The results shown in Fig. 13 might suggest that we should use different gains for different applications. We considered this approach but found it to have two practical problems.

First, as shown in Fig. 2, measurement error depends not only upon the type of data but also upon load size. And, although we could allow the gains to be a function of load size and data type, multiple versions of SPA would be relatively more difficult to implement and maintain.

Second, and more important, in practice it will not be possible to determine the exact value of G. For, in general, actual loads will not display constant incremental growth, but may exhibit random fluctuations about a trend line. In this case, the model described in Section III is still appropriate if x_n denotes the trend level, instead of the true load. However, σ^2 now consists of two components: one because of the difference between the measured and true loads and the other because of the difference between the true load and the trend line. Although

Ref. 2 quantifies the first component, we have no way a priori to assess the contribution of the second component. Thus, our assumed value for σ^2 may differ from the actual value.

Similarly, since we cannot determine a priori the difference between our estimate of the aggregate and individual trunk group load growth rates, our estimate of σ_g^2, eq. (19), will, in general, differ from the actual value.

Consequently, the assumed value of G will in general also differ from the actual value. Accordingly, to guard against the consequences of an error in our estimate of G, we would like to employ an algorithm which performs well under a variety of possible operating conditions.

4.2.3 Robust gains for linear growth

Fortunately, there exists a robust set of gains which provides the same average performance independent of the true value of G. That is, as shown by the curve labeled "robust" in Fig. 13, if we use the gains corresponding to $G = G_3$ (see Fig. 12) then the 5-year average rms error will be about 10 percent less than that of the conventional projection method—independent of the actual value of G.

Of course, as with any robust technique, Fig. 13 shows that we pay a premium by receiving less than the theoretically optimal performance for protection against the possibility of a performance substantially worse than that of the conventional projection method. However, we used the data from our field test[5] to estimate, a posteriori, the value

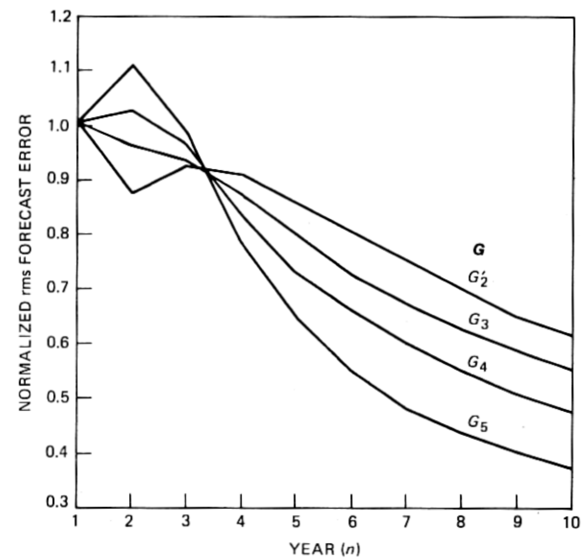

Fig. 14—Robust filter (variable gains): rms forecast error vs. year.

of G. Remarkably, the observed value was $G \simeq G_3$. Thus, at least for this one case, our robust filter was in fact nearly optimal.

Although the robust gains yield an average performance which is independent of G, Fig. 14 shows that the actual performance in each year is not independent of G. For example, if $G = G_5$ the rms error of SPA increases above that of the conventional method by about 10 percent in year 2, but is decreased by about 35 percent in year 5. Thus, during the first couple of years after initialization, SPA may perform slightly worse than the conventional method for some trunk groups. Thereafter, SPA will perform better, provided that the trend remains constant.

4.3 Robust, constant gains for SPA

4.3.1 Response to unexpected changes in trend

With the "robust gains for linear growth," SPA is robust to measurement and initial growth estimation errors. For practical applications, however, it is equally important that SPA also be made robust to deviations from linear growth.

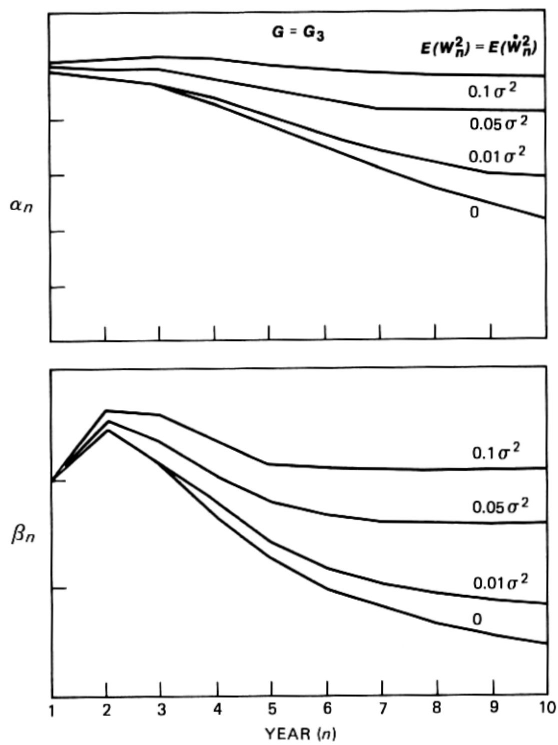

Fig. 15—Optimal gains with modelling error.

That is, the results of Section 4.2.3 assume that the trend line displays constant incremental growth. Under this assumption, since the forecast error approaches zero, the gains α_n and β_n approach zero as n increases. Accordingly, as discussed in Section 2.3, SPA would eventually respond to unexpected changes in the level or slope of the trend only when their cumulative effect produced two consecutive outliers of the same sign.

4.3.2 Constant gains

At the expense of receiving less than the theoretically optimal performance in the ideal case where the trend remains constant, we can decrease SPA's response time to unexpected changes in trend by not allowing the gains α_n and β_n to approach zero. As discussed in Ref. 7, one approach to selecting limiting values for α_n and β_n is to add zero-mean, uncorrelated random variables (w_n and $\dot{w}_n$) to the description of the true load in eq. (9). That is, we assume that the covariance matrix Q of modelling errors is nonzero. These modelling error terms lead to gains α_n and β_n which approach nonzero constants that depend upon the mean square value of these error terms. For example, Fig. 15 shows the optimal gains corresponding to $G = G_3$ for several values of $E(w_n^2)$ and $E(\dot{w}_n^2)$.

Although theoretically appealing, the above approach leaves the practical problem of determining appropriate values for $E(w_n^2)$ and $E(\dot{w}_n^2)$. Values could be gleaned from a large amount of historical data,

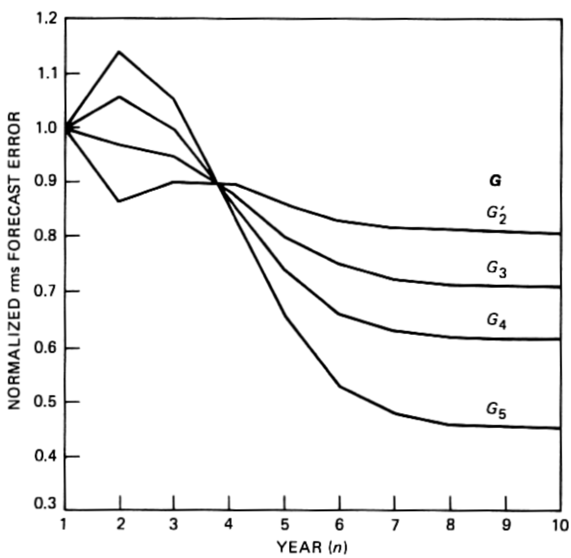

Fig. 16—Robust filter (constant gains): rms forecast error vs. year.

which we do not have, but there is no guarantee that they would be appropriate for the future.

Therefore, instead of pursuing the above approach, we simply replaced the variable gains corresponding to $G = G_3$ in Fig. 12 by the average of the first few terms and made the following observations: First, as shown in Fig. 16, the performance in the ideal case where the trend remains constant is nearly identical for the first six years after initialization to that for the robust gains for linear growth. Thereafter, the performance is somewhat poorer, but we anticipate that only a small fraction of groups will remain on a constant linear trend beyond 6 years. Second, as illustrated in Fig. 17, when the true load displays random deviations from a linear trend, i.e., when w_n, $\dot{w}_n$ are nonzero, the constant gains perform better than those designed for linear growth. Thus, although the performance is somewhat degraded in the ideal case, the use of constant gains provides protection against the very real possibility of unexpected changes in trend. For comparison, Fig. 17 also displays the performance with gains obtained by truncating the robust gains for linear growth in year 6. As indicated, performance is somewhat better in the ideal case, but worse when modelling errors are present. Since we do not have sufficient historical data to estimate the level of modelling errors, we are, therefore, recommending the use of the constant gains.

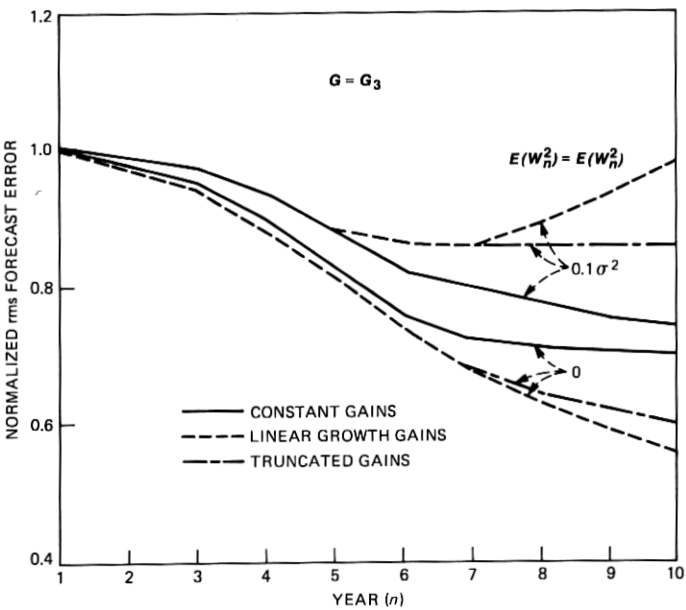

Fig. 17—Forecast error with modelling error.

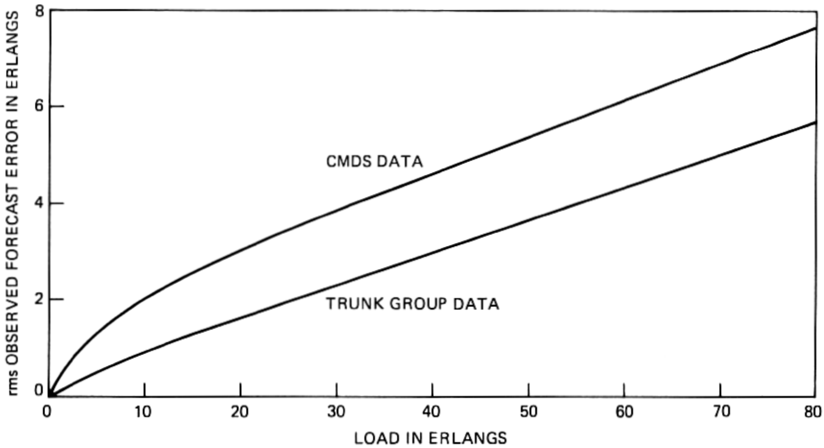

Fig. 18—Root-mean-square observed forecast error vs. load: outlier thresholds.

V. OUTLIER THRESHOLDS

As discussed in Section 2.3, Huber's studies[5] show that the outlier thresholds may be set anywhere in the range of one to two times the rms observed forecast error. In theory, the observed forecast error depends upon the number of data points processed after initialization; however, since the average rms forecast error decreases by only about 10 percent and since performance is not sensitive to the precise setting of the thresholds, we will set them based upon the initial value of the rms observed forecast error; that is, by

$$\rho^2 = E(x_{1,0} - y_1)^2. \tag{22}$$

Since $x_{1,0} = (1 + \hat{g})y_0$ and $x_1 = (1 + g)x_0$, it follows that

$$\begin{aligned}\rho^2 &= E[(y_0 - x_0) + (y_0 - x_0)\hat{g} + x_0(\hat{g} - g) - (y_1 - x_1)]^2 \\ &= 2\sigma^2(1 + \hat{g} + \hat{g}^2/2) + s_0^2\sigma_g^2,\end{aligned} \tag{23}$$

or, since $|\hat{g}| \ll 1$,

$$\rho^2 \approx x_0^2\sigma_g^2 + 2\sigma^2. \tag{24}$$

To establish numerical values for ρ, we use the theoretical expressions for mean square measurement error given in Ref. 2:

$$\sigma^2 = \frac{1}{20}\left(\frac{2x_0h}{p} + V_d\right), \tag{25}$$

where h is the mean call-holding-time in hours, p is the sampling rate (for trunk group data, $p = 1.0$; for CMDS data, $p = 0.05$), and

$$V_d = \max(0,\, 0.13x_0^{\phi} - 2x_0h) \tag{26}$$

is the variance of the daily source loads. For first-routed loads, $\phi = 1.5$ is usually appropriate.

Figure 18 displays ρ as a function of the load x_0 for both trunk group data ($p = 1.0$) and CMDS data ($p = 0.05$). In each case, we assume $\sigma_g = 0.06$, which is the value observed during our field test,[6] and $h = 1/12$.

Based upon the results of Fig. 18 and Huber's studies,[5] it follows that outlier thresholds set at two times the value of ρ for trunk group data should provide adequate performance for both trunk group and CMDS data. That is, this setting places the thresholds within the allowed range of about one to two times the actual value of ρ for both trunk group and CMDS data.

VI. CONCLUSIONS

The SPA employs a linear two-state Kalman filter, together with logic for detecting and responding to outlier measurements. The parameters of SPA have been selected to provide improved performance over the range of operating conditions, including the use of either trunk group or point-to-point traffic measurements.

Under the assumption of linear growth for 5-year intervals, the average rms 1-year forecast of SPA is about 10 percent less than that of the forecasting methods currently in use in the Bell System. Moreover, the filter parameters and outlier procedures have been designed so that SPA will respond to changes in trend.

Field test results confirm the theoretical results presented here. Accordingly, specifications have been written for the inclusion of SPA in the Bell System's standard mechanized trunk forecasting systems.

REFERENCES

1. A. J. David and C. D. Pack, "The Sequential Projection Algorithm: A New and Improved Traffic Forecasting Procedure," International Teletraffic Congress 9, Malaga, Spain, 1979.
2. J. P. Moreland, "Estimation of Point-to-Point Telephone Traffic," B.S.T.J., *57,* No. 8 (October 1978), pp. 2857–63.
3. R. L. Franks et al., "A Model Relating Measurement and Forecast Errors to the Provisioning of Direct Final Trunk Groups," B.S.T.J., *58,* No. 2 (February 1979), pp. 351–78.
4. C. J. Masreliez and R. D. Martin, "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," IEEE Trans. Automatic Control (June 1977) pp. 361–71.
5. P. J. Huber, "Robust Estimation of a Location Parameter," Annals of Mathematical Statistics, *35* (1964), pp. 73–110.
6. R. H. Harris and R. E. Sharkey, "Implementing the Sequential Projection Algorithm in an Operating Telephone Company," International Teletraffic Congress 9, Malaga, Spain, 1979.
7. C. D. Pack and B. A. Whitaker, "Kalman Filter Models for Network Forecasting," B.S.T.J., this issue.