THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Volume 61 | January 1982 | Number 1

Kalman Filter Models for Network Forecasting

By C. D. PACK and B. A. WHITAKER

(Manuscript received December 31, 1980)

The Bell System has recently completed studies that are expected to result in substantially improved forecasts for use in network planning. These improved forecasts are achieved through the use of new forecasting algorithms that employ Kalman filter models. To motivate the selection of Kalman filter forecasting procedures, we describe the Bell System's special data characteristics and processing requirements in the network planning process. We also discuss the Kalman filter models, their statistical properties, the model identification process, and certain implementation considerations.

I. INTRODUCTION

Projections of message circuit usage (as measured in hundred call seconds or CCS), from which message circuit requirements (trunks) are developed, and special services* circuit demand are fundamental parts of the Bell System's network planning and provisioning process. An overview of the information flows in this process is shown in Fig. 1; more details are given in Refs. 1 to 4 and in the companion forecasting papers in this issue.

Since these projections strongly influence the allocations of several billions of construction dollars annually, it is important that they possess high quality statistical properties. For example, the projections should be unbiased; that is, the forecasts should not be consistently

* *Special services* is a generic term referring to all Bell System services other than ordinary message telephone service. Examples include foreign exchange lines, tie lines, and private lines.

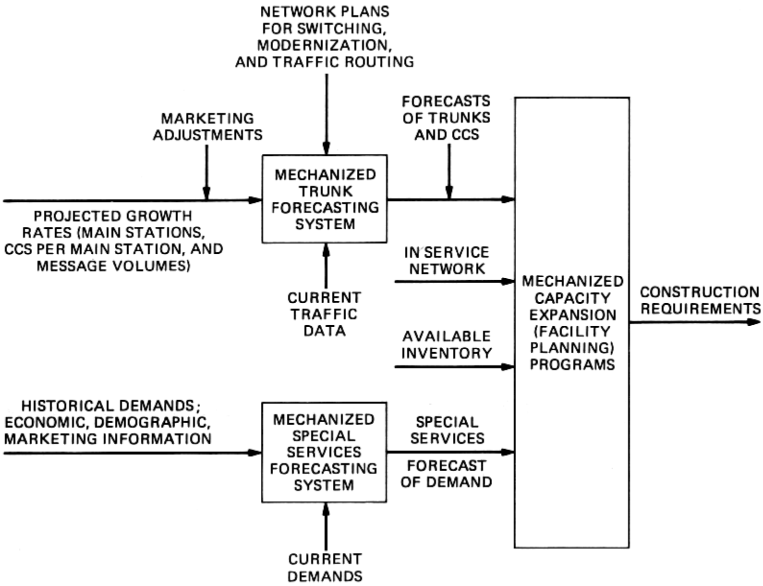

Fig. 1—Network planning and provisioning process.

high or low. A biased forecast will result in either over expenditures or equipment shortages, depending on the sign of the bias. Also, the forecasts should be stable, or precise, never varying too much from the realized true requirements. Previous studies have shown that highly variable forecasts result in increased reserve capacity requirements necessary to meet customer demands.[5]

The service and economic motivations for high quality forecasts of message trunks and special services circuits have led the Bell System to reevaluate the existing projection algorithms used in these processes and to recommend improved methods, as necessary. It should be noted that a further motivation for this reevaluation is that most existing projection methods utilize curve fitting and extrapolation algorithms that were originally designed for manual calculations and that were available prior to the widespread use of computers and the advent of modern estimation techniques.[1,2]

The possible approaches to improved projection methods were influenced strongly by the particular characteristics of the time series and by the typical mode in which the forecasts are produced.

Two factors were crucial in the selection of forecasting models and algorithms: the need to produce a large number of forecasts in a fully mechanized system and the typical dearth of data for any individual series.

A large number of forecasts is necessary because trunk and special services circuit requirements must be forecast at least once per year for each trunk group and special services circuit group on record. Thus, there may be up to 100,000 trunk group time series and tens of thousands of special services circuit time series forecast in each run of the mechanized forecasting systems illustrated in Fig. 1. Therefore, these procedures must be fully mechanized. Moreover, for ease of comprehension and for computational efficiency, the algorithms must be simple to program and maintain.

The sparsity of data is perhaps the more restrictive of the two key considerations in selecting new projection algorithms. Typically, the individual time series have between 1 and 10 years of relevant data, with 3 to 5 years being common. The fact that a projection may have to be based on less than 5 years of data (up to 60 points if monthly data were available), make the use of many approaches, including the popular Box-Jenkins ARIMA (autoregressive integrated moving average) models[6] that require more than 100 data points, infeasible.

The linear Kalman filter, in its most general form, was derived by Kalman[7] and Kalman and Bucy[8] in the early 1960s. The motivation for their work had its roots in various control theory applications. Of particular note was the implementation of the algorithms in tracking systems for the aerospace industry. In recent years, the models have also been used in power systems, process control, and forecasting.

In Section III, we describe the general Kalman filter linear model, analyze and interpret the key matrices of the formulation, and relate the Kalman filter to other common forecasting models. In Section IV, we discuss various implementation considerations that are necessary to reduce the Kalman filter algorithm to practice. In Section V, we indicate methods for evaluating the effectiveness of the Kalman filter models and discuss the importance of certain key statistics. Finally, in Section VI, we summarize our conclusions concerning the use of Kalman filters for forecasting in the Bell System.

Three companion papers to this overview follow. These papers describe the successful use of Kalman filter models for three Bell System forecasting applications—two concern message trunk forecasting and one covers special services demand forecasting.

The first trunk forecasting paper, by J. P. Moreland, describes a simple linear, two-state model for use in projecting busy season (yearly peak) trunk group loads.[9] The second paper, by A. Ionescu-Graff, describes a linear, two-state Kalman filter with an absorbing barrier that can improve the quality of special services demand forecasts.[11] The third paper, by C. R. Szelag, derives a Kalman filter for traffic that has not only linear growth, but also a seasonal within-year pattern as well.[10] That is, Szelag illustrates how, for trunk group load patterns exhibiting seasonality, nonbusy season data can be used to update and improve estimates of imminent busy season loads.

II. KALMAN FILTER DESCRIPTION

The Kalman filter, described in more detail in Section III, has many desirable properties. Most of these properties are not unique to the Kalman filter; however, because of its generality and particular form, as well as statistical properties, computational characteristics, and robust qualities it should be considered in most estimation applications.

2.1 Models

The models used are based on state-space representations of the variables being estimated. The state-space formulation implies that, at each point in time, the process being modeled is described by a vector of state variables that summarize all relevant quantities of interest. In most instances, the state variables have physical interpretations, such as trunk quantities, growth rates, and so on. Then, a further characterization of the model specifies how this state vector evolves over time.

The Kalman filter algorithm uses this model of the time behavior of the system along with "noisy" observations or measurements of some system variables to produce optimal estimates of all state variables. These estimates are then used in the process model to determine state estimates for future time periods.

The distinction between the state-space models and the ARIMA models is mostly in the model identification process. That is, while the interpretation of the state-space models permits the user to choose a model based on physical properties (and hence when only limited data are available), the "time series" approach of Box and Jenkins attempts to have the data specify the model based on certain first- and second-order characteristics. However, after a state-space time-series model is known, one can find nearly equivalent representations of that model for either theory.

The particular state-space formulation of Kalman has some desirable features. First, it lends itself to simple, recursive estimation of the parameters. That is, no data history need be stored. As new data or observations become available, they are processed and the stored state vectors are updated accordingly. In fact, this recursive calculation suggests a Markovian property of the filter: the current state vector summarizes all relevant historical information concerning the history of the time series. Second, there is a provision for separate characterization of the two sources of significant estimation errors: the dynamics of the true process and relationship between the state variables and the measurements used to estimate these variables. Third, the model provides an analytic framework for studying relationships among the first- and second-order properties of the state variables and measurements. For example, one can derive analytic expressions for forecast variances as functions of the number of data points processed and the autocorrelation matrix of the measurement errors. Finally, the model is general enough to include as special cases the common models: exponential smoothing, weighted least squares, multiple linear regression, and Wiener filtering.

2.2 Statistical properties

A correctly specified linear Kalman filter produces forecasts that have minimum mean square error.* Moreover, the forecasts are unbiased and have minimum variance. When the errors are Gaussian, these properties hold without restriction to the class of linear models. The estimators can also be derived using maximum likelihood or Bayes models. When the models are only approximately correct, the generality of the formulation allows one to analyze the filter's suboptimal performance and, if desired, to adjust the filter's parameters as appropriate.[12,13]

III. DISCRETE-TIME LINEAR KALMAN FILTER MODEL

3.1 The model

It is assumed that the true process dynamics are described by the following linear transition equation

$$\mathbf{X}_{n+1} = \phi \mathbf{X}_n + \mathbf{U}_n + \omega_n \, , \tag{1}$$

where

$\mathbf{X}_n$ = an s-vector of state variables in period n,

ϕ = an $s \times s$ transition matrix that may, in general, depend on n,

ω_n = an s-vector of random modeling errors, i.e., random deviations of the true process from the assumed linear relation defined by ϕ, and

$\mathbf{U}_n$ = an s-vector of deterministic changes in state.

The one-step projection formula is given by

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

where, in general, $\hat{\mathbf{X}}_{n+k,n}$ is an estimate of $\mathbf{X}_{n+k}$ $(k \geq 0)$ given data $\mathbf{y}_1, \cdots, \mathbf{y}_n$ in periods 1 through n, where $\mathbf{y}_n$ is a d-vector of observed variables in period n.

The relations that distinguish the Kalman filter model and associated computational procedures from other linear estimation techniques are the particular model relating $\mathbf{X}_n$ to $\mathbf{y}_n$ and the algorithm for

* These terms and others are defined in Section V.

computing $\hat{\mathbf{X}}_{n,n}$. The $d \times s$ matrix H, which in general may depend on n, defines the relationship between $\mathbf{y}_n$ and $\mathbf{X}_n$ as

$$\mathbf{y}_n = H\mathbf{X}_n + \boldsymbol{\nu}_n , \tag{3}$$

where $\boldsymbol{\nu}_n$ is a d-vector of measurement errors. At time n, the vector $\hat{\mathbf{X}}_{n,n}$ is computed by

$$\hat{\mathbf{X}}_{n,n} = \hat{\mathbf{X}}_{n,n-1} + K_n(\mathbf{y}_n - H\hat{\mathbf{X}}_{n,n-1}). \tag{4}$$

The $s \times d$ "Kalman gain" matrix K_n can be calculated recursively by the following equations:

$$\begin{aligned} K_n &= P_nH^T(HP_nH^T + R)^{-1} \\ S_n &= (I - K_nH)P_n \\ P_{n+1} &= \phi S_n \phi^T + Q , \end{aligned} \tag{5}$$

where

(*i*) R is the covariance matrix of the measurement errors, i.e., $R \equiv E(\boldsymbol{\nu}_n\boldsymbol{\nu}_n^T)$,

(*ii*) Q is the covariance matrix of the modeling errors, i.e., $Q \equiv E(\omega_n\omega_n^T)$, and

(*iii*) it is assumed that $E(\boldsymbol{\nu}_n) = E(\omega_n) = 0$ for all n, $E(\omega_n\boldsymbol{\nu}_i^T) = 0$ for all (n, i), $E(\omega_n\omega_i^T) = 0$ for all $n \neq i$, and $E(\boldsymbol{\nu}_n\boldsymbol{\nu}_i^T) = 0$ for $n \neq i$, and

(*iv*) in general, Q and R may depend on n.

Thus, in summary, the forecasting procedure has the following steps:

1. When $n = 0$, the filter is initialized by a user-supplied estimate $\hat{\mathbf{X}}_{0,0}$ of the initial state vector $\mathbf{X}_0$ and S_0.
2. Using these estimates, a one-period-ahead forecast is produced using eq. (2).
3. The gain matrix, K_n, and the matrices S_n and P_{n+1} are calculated.
4. When a new observation is received, eq. (4) is used to obtain a "smoothed" estimate $\hat{\mathbf{X}}_{n,n}$ of the present state vector.
5. Then using this new estimate, a one-period-ahead forecast is produced. The period index n is incremented and Step 3 is repeated.

The algorithm continues to process new observations and produce forecasts in this manner.

Several points should be noted:

1. If the matrices H, R, ϕ, and Q are independent of n, the gain matrix, K_n, and the matrices S and P are independent of the observations. Thus, these matrices can be precalculated for use in the algorithm.
2. No past observations must be stored since all historical information is contained in the "smoothed" estimate $\hat{\mathbf{X}}_{n,n}$ or, equivalently via eq. (2), the one-period-ahead projection $\hat{\mathbf{X}}_{n+1,n}$.

3. Only the one-period-ahead forecast must be saved to be used in the next period's process.

When the assumptions listed above in (*iii*) are valid and the models accurately describe the true process dynamics and the measurement system, the Kalman filter produces unbiased estimates of $\hat{\mathbf{X}}_{n+k,n}$; that is, $E(\hat{\mathbf{X}}_{n+k,n}) = \mathbf{X}_{n+k}$ for $k \geq 0$. In addition, the estimates $\hat{\mathbf{X}}_{n+k,n}$ have minimum variance in the class of all unbiased estimators.[13] If ω_n and ν_n are Gaussian, then no restriction to the class of linear estimates is required and the estimators can also be derived from maximum likelihood and Bayesian models.[13] Conveniently, the Kalman formulation actually provides estimators for the estimation error variance matrices of interest:
(smooth)

$$S_n \equiv E(\hat{\mathbf{X}}_{n,n} - \mathbf{X}_n)(\hat{\mathbf{X}}_{n,n} - \mathbf{X}_n)^T, \tag{6}$$

and, for $k \geq 1$,
(predict)

$$P_{n+k} = E(\hat{\mathbf{X}}_{n+k,n} - \mathbf{X}_{n+k})(\hat{\mathbf{X}}_{n+k,n} - \mathbf{X}_{n+k})^T. \tag{7}$$

Therefore, the requirement that the user specify an initial estimate of S_0 is a need for an estimate of the covariance matrix of the initial state vector $\hat{\mathbf{X}}_{0,0}$.

Analyses of the sensitivity of the filter performance to model accuracy and of the modified estimators of error covariance matrices for the case of suboptimal gains K' is given in Ref. 13.

3.2 Interpretation

From eq. (4), we see that the vector $\hat{\mathbf{X}}_{n,n}$, the smoothed estimate of $\mathbf{X}_n$, is derived as the previous one-step projection $\hat{\mathbf{X}}_{n,n-1}$, plus a linear combination (weighting) of the differences between the measurement $\mathbf{y}_n$ and the previous estimate (forecast) of these measurements, $H\hat{\mathbf{X}}_{n,n-1}$. The weights assigned to the difference terms are appropriate components of the gain matrix K_n. It is also important to note that $\hat{\mathbf{X}}_{n,n}$ depends on $\hat{\mathbf{X}}_{n,n-1}$, $\mathbf{y}_n$, K_n, and H, but not explicitly on $\mathbf{y}_1, \cdots, \mathbf{y}_{n-1}$. This recursive nature of the Kalman filter eliminates the need for storage of historical data.

We can give some insight into the effect K_n has on algorithm performance, without actually describing a specific model. It can be seen from eq. (5) that K_n has terms which are directly proportional to the elements of the covariance matrix Q* and inversely proportional

* As we indicate later in Section 3.4, this "proportionality" to Q is strongest for large n.

to the elements of R. That is, K_n is, in a sense, proportional to the variability of the true process dynamics and inversely proportional to the measurement variability. Thus, it is the "Q/R" relationship which defines the responsiveness of the filter (via the K matrix) to estimated errors in state $(\mathbf{y}_n - H\hat{\mathbf{X}}_{n,n-1})$. As the elements of K_n decrease (by decreasing Q or increasing R), the forecasts become more stable. That is, if we have more confidence in the unbiasedness of the process model or less confidence in our observations, the filter will be designed to respond more slowly to apparent deviations from the predicted trend line. As K_n increases (by increasing Q or decreasing R), forecasts detect and respond to data that deviate from the assumed deterministic linear model, represented by the expected value of eq. (1).

3.3 An example

Consider a simple linear two-state model ($s = 2$), where

$$\mathbf{X}_{n+1} = \begin{bmatrix} x_{n+1} \\ \dot{x}_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ \dot{x}_n \end{bmatrix} + \begin{bmatrix} \omega_n \\ \dot{\omega}_n \end{bmatrix} \tag{8}$$

and ($d = 1$)

$$y_n = x_n + \nu_n . \tag{9}$$

We further assume that $R \equiv 1$ (since the gain matrix depends only on Q/R, we do this with no loss in generality). To complete the model, the 2×2 matrix Q must be specified. This simple two-state model is fundamental to the models presented in the three companion papers. In these models, x_n is referred to as the level of the process, and $\dot{x}_n$, as the growth increment.

We now show how the various components of Q influence the response of the filter to a measurement y_n. First, note that eq. (4) can be rewritten as

$$\hat{\mathbf{X}}_{n,n} = \begin{bmatrix} \hat{x}_{n,n} \\ \hat{\dot{x}}_{n,n} \end{bmatrix} = \begin{bmatrix} \hat{x}_{n,n-1} + k_{11}^{(n)}(y_n - \hat{x}_{n,n-1}) \\ \hat{\dot{x}}_{n,n-1} + k_{21}^{(n)}(y_n - \hat{x}_{n,n-1}) \end{bmatrix}. \tag{10}$$

Hence, the smoothing process is determined by the specification of two sequences of number $k_{11}^{(1)}, \cdots, k_{11}^{(n)}, \cdots$ and $k_{21}^{(1)}, \cdots, k_{21}^{(n)}, \cdots$. As we indicated above, these gains tend to be proportional to the elements of Q.

Fig. 2 shows the filter operation and the role of the gain sequence $\{k_{11}^{(n)}\}$. Initially, attention should be focused on the trend line at time $n - 1$ derived from $n - 1$ pieces of data, $y_1, \cdots, y_{n-1}$ and note that $\hat{x}_{n-1,n-1}$ lies on that line. Further, $\hat{x}_{n,n-1}$ is a straight projection of this trend one-period ahead. When the new observation, y_n, is obtained, the slope of the trend line is adjusted upward in the direction of y_n

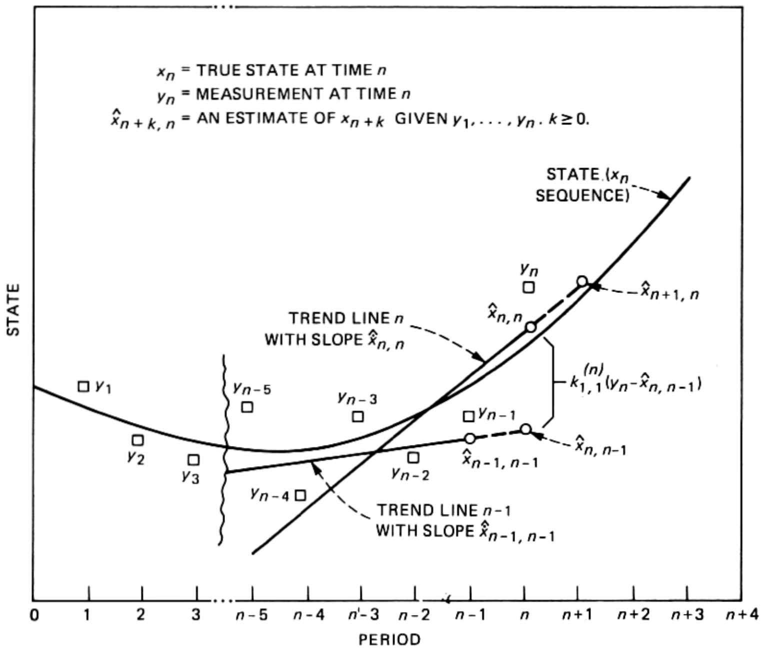

Fig. 2—Projection using the Kalman filter.

relative to $\hat{x}_{n,n-1}$. The smoothed estimate $\hat{x}_{n,n}$ of x_n is $\hat{x}_{n,n-1}$, plus a factor $k_{11}^{(n)}$ times the increment $y_n - \hat{x}_{n,n-1}$. The new trend line passes through $\hat{x}_{n,n}$ and its extrapolation produces $\hat{x}_{n+1,n}$.

The choice of Q (and R) determines the $k_{11}^{(n)}$ and, hence, the response of the filter to observation y_n relative to $\hat{x}_{n,n-1}$. Note that the *smoothed* estimate $\hat{x}_{n,n}$ of x_n, *not* y_n, is projected into the future. Thus, we do not project all of the noise in y_n into the future. This is a fundamental distinction between Kalman filter projections and most existing load-projection algorithms used by the Bell System.

3.4 Estimating matrices

If all Kalman filter model matrices (ϕ, H, S_0, Q, and R) are known, the algorithm is completely specified; moreover, the desirable statistical properties of the filter are assured. However, in general, the true model is not known precisely, or it is likely that the model is to be applied to many different estimation problems. Hence, in practice, we often settle for less optimal (in a bias or variance sense), but more robust properties for our forecasts. That is, we choose to achieve a reasonable balance between bias and variability (stability) over a wide range of practical interest for key parameters while relaxing some of the assumptions in Section 3.1.

Usually, it is the case that the processes being modeled suggest a reasonable choice of matrices ϕ and H. (Examples are discussed in the companion papers in this issue.) Also, typically, the data characteristics can be analyzed so that R is known either analytically or empirically. Therefore, the main concern is the selection of matrices S_0 and Q. The former strongly influences the transient characteristics of the filter; that is, K_n is strongly dependent on "S_0/R" for n small (<10). The latter affects the steady-state performance; that is, K_n approaches "Q/R" for large n.

These matrices S_0 and Q can be determined analytically or empirically to provide the desired responsiveness and statistical properties over many time series. The key to the analytical approach is the matrix P_{n+1} in eq. (5), which describes forecast variance as a function of model matrices, when the assumptions of Section 3.1 are correct. Empirical studies attempt to tune filter performance so that a robust balance of unbiasedness and stability is achieved over the test cases. The performance trade-offs are illustrated generically in Fig. 3 as a function of K_n whose dependence on S_0 and Q was stated previously. The studies described in the companion papers illustrate these two approaches to filter design.

We have ignored the extensive literature (see, for example, Ref. 14) on model identification for Kalman filters, because, as for the Box-Jenkins technique, substantial data is required for each time series.

3.5 Special cases

The Kalman filter model includes as special cases many common estimation techniques. We indicate three such examples.

3.5.1 Multiple regression

The multiple regression approach to estimation assumes that a time series $\{y_n\}$ is well-approximated by

$$y_n = H_n^T \mathbf{X} + \nu_n , \tag{11}$$

where H_n is an s-vector of observations of the independent regression variables, $\mathbf{X}$ is an s-vector of (constant) regression weights, and ν_n is an error term. The Kalman filter model is obtained by allowing the regression coefficients to depend on n and by adding the dynamics relation

$$\mathbf{X}_{n+1} = \mathbf{X}_n . \tag{12}$$

Clearly, the model includes the autoregressive case (H_n is composed of previous realizations of y_n) and can be generalized to vectors $\mathbf{y}_n$.

3.5.2 Exponential smoothing

Exponential smoothing is a process whereby current estimates of

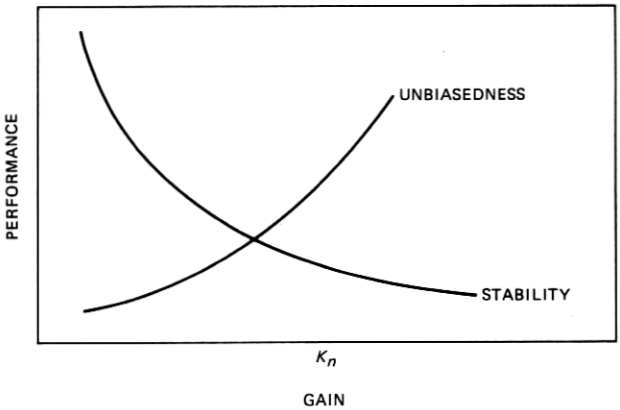

Fig. 3—Empirical performance trade-offs.

state $\hat{x}_{n,n}$ are composed of a weighted average of the previous estimate at state ($\hat{x}_{n-1}$) and the current measurement (y_n). That is,

$$\hat{x}_n = \hat{x}_{n-1} + k(y_n - \hat{x}_{n-1}). \tag{13}$$

Note that the weight k is a constant. The term exponential smoothing was chosen because the previous measurements influence current estimates by weights that decrease exponentially with lag.

Clearly, eq. (13) is of the form of eq. (4) and can be generalized to a Kalman filter by including eqs. (2) and (3), and the dependence of k on n.

3.5.3 Wiener filtering

Wiener filtering[15] corresponds to the case of constant Kalman gain matrix, i.e., when $K_n \equiv K$ for all n. If ϕ, H, R, and Q are constant, then the steady-state error covariance matrices exist and a Kalman filter with gains calculated using these matrices is identical to a Wiener filter.

IV. IMPLEMENTATION CONSIDERATIONS

Experience has shown that the Kalman algorithm, as described in Sections II and III, performs well, as long as the model is reasonable and the data $\mathbf{y}_n$ are consistent with the model assumptions. However, in practice, additional logic or considerations are necessary to improve the performance when outlier data are present or when certain types of nonstationarity are present and to improve the computational efficiency of the procedures.

An overview of a typical Kalman filter implementation is shown in Fig. 4. We will describe the components of the Kalman filter system in the following subsections.

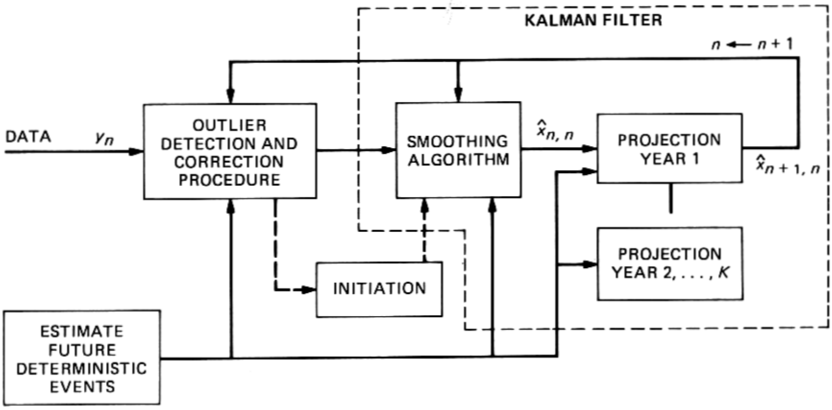

Fig. 4—Kalman filter implementation.

4.1 *Data smoothing and projection*

The Kalman filter algorithm, described in Section 3.1, comprises the smoothing and projection functions.

4.2 *Initiation, transient response, and gain sequences*

In some Kalman filter applications, the transient period is very short relative to the total time the filter is in operation. However, in the applications described in Refs. 9 to 11, a series rarely exists for more than 10 years, with 3 to 5 years being typical. Therefore, the transient response of the filter is important and $\hat{\mathbf{X}}_{0,0}$ and S_0 (and, hence, K_n for small n) must be carefully chosen to provide good performance during the transient period.

Surprisingly, good statistical performance in both the transient and equilibrium states is often achieved with fixed ($K_n = K$ for all n) or finite ($K = K_{n*}$ for $n \geq n^*$) gain sequences. Moreover, significant computational efficiencies are obtained when the precomputed finite gains are stored for on-line use. The companion papers describe the success of fixed and finite gain sequences for the respective applications.

4.3 *Outlier detection, deterministic events*

We define unusual data to be that representing either an unusually large (outlier) measurement error ($\boldsymbol{\nu}_n$) or an unforeseen deterministic event ($\mathbf{U}_n$). The importance of these two types of errors is that the smoothing process (4) tends to adjust the previous forecast $\hat{\mathbf{X}}_{n,n-1}$ in the direction of the new measurement $\mathbf{y}_n$, with the movement being proportional to the estimated error ($\mathbf{y}_n - H\hat{\mathbf{X}}_{n,n-1}$). Therefore, without additional logic of an adaptive nature, outliers would cause overreaction to measurements; unforeseen deterministic events would be insufficiently accounted for because the *full* (rather than smoothed) impact should have been entered in eq. (2).

However, in practice, one cannot always satisfy the two conflicting objectives—ignore bad measurements and react strongly to unforeseen deterministic events. Hence, either robust estimators must be employed or logic must be provided to identify and distinguish the two cases and to not overreact to either. The balance is again between bias and stability: over-response reduces bias but degrades stability, and vice versa.

Considerable discussion of this trade-off and the resulting filter design for one application is described in Ref. 9.

V. MODEL EVALUATION

In previous sections of this paper, we have alluded to various statistical criteria that may be used to evaluate the performance of the filter. Clearly, none is right or wrong unless it can be shown that some significant costs are directly and uniquely related to a particular criterion. We have never seen such justifications. However, it is usually the case that costs are related in some indirect fashion to various first- and second-order error statistics of the forecasts. The companion papers will each refer to some subset of the following statistics, possibly normalized to be stated as a percent:

Bias

$$A_{n+k,n} \equiv E(\hat{\mathbf{X}}_{n+k,n} - \mathbf{X}_n)$$

(In)stability (view-over-view variability)

$$S_{n+k,n} = E(\hat{\mathbf{X}}_{n+k,n} - \hat{\mathbf{X}}_{n+k,n-1})(\hat{\mathbf{X}}_{n+k,n} - \hat{\mathbf{X}}_{n+k,n-1})^T$$

(Im)precision or variance

$$P_{n+k,n} = E(\hat{\mathbf{X}}_{n+k,n} - E(\hat{\mathbf{X}}_{n+k,n}))(\hat{\mathbf{X}}_{n+k,n} - E(\hat{\mathbf{X}}_{n+k,n}))^T$$

Mean square error

$$M_{n+k,n} = E(\hat{\mathbf{X}}_{n+k,n} - \mathbf{X}_n)(\hat{\mathbf{X}}_{n+k,n} - \mathbf{X}_n)^T$$

Clearly, there exist some analytical relationships among these statistics. We will not derive any here; however, it is of interest to point out that, if $\hat{\mathbf{X}}_{n+k,n}$ and $\mathbf{X}_n$ are uncorrelated (a sufficient condition is that $Q \equiv 0$ or $k = 0$ for the linear model of Section III), then

$$M_{n+k,n} = A_{n+k,n}A^T_{n+k,n} + P_{n+k,n}$$

which, in the scalar case, reduces to

$$M_{n+k,n} = A^2_{n+k,n} + P_{n+k,n} .$$

The rms error is commonly calculated as $(M_{n+k,n})^{1/2}$ for this last case.

VI. STATUS AND CONCLUSIONS

The Kalman filter model promises to provide forecasts, for use in Bell System network planning, that are both substantially improved in a statistical sense relative to existing methods and computationally more efficient. The former claim is based on testing by analysis, computer simulation, and field study. The latter, because of the efficient recursive calculations of the filter, has been borne out through Bell Laboratories programs and limited field studies. The companion papers in this issue will elaborate on these conclusions and will describe plans to implement the algorithms in standard, mechanized network planning tools.

REFERENCES

1. N. D. Blair, "Forecasting Telephone Traffic in the Bell System," International Teletraffic Congress 8, Melbourne, November 1976, pp. 211:1–11.
2. W. C. Johnson and W. P. McGuire, "The Trunk Forecasting System Tells What, Where, When," Bell Laboratories Record (June 1974), pp. 192–8.
3. C. Cirillo, "Controlling Circit Orders with TIRKS," Bell Laboratories Record (March 1977), pp. 73–7.
4. R. C. Chang, R. D. Pease, and G. L. Wright, "Facility Planning. . .Over the Long Haul It's IFRPS," Bell Laboratories Record (June 1980), pp. 196–203.
5. 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* (February 1979), pp. 351–78.
6. G. E. P. Box and G. M. Jenkins, *Time Series Analysis*, San Francisco: Holden-Day, 1970.
7. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., *82* (1960), pp. 340–5.
8. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Trans. ASME Ser. D, J. Basic Eng., *83* (1961), pp. 95–107.
9. J. P. Moreland, "A Robust Sequential Projection Algorithm for Traffic Load Forecasting," B.S.T.J., this issue.
10. C. R. Szelag, "A Short-Term Forecasting Algorithm for Trunk Demand Servicing," B.S.T.J., this issue.
11. A. Ionescu-Graff, "A Sequential Projection Algorithm for Special-Services Demand," B.S.T.J., this issue.
12. R. K. Mehra, "Kalman Filters and their Applications to Forecasting," *TIMS Studies in the Management Sciences*, Amsterdam: North-Holland Publishing Co., *12*, 1979, pp. 75–94.
13. A. Gelb, ed., *Applied Optimal Estimation*, Cambridge, Mass.: MIT Press, 1974.
14. R. K. Mehra and D. G. Laintootis, eds. *System Identification, Advances and Case Studies*, New York: Academic Press, 1976.
15. N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications*, New York: John Wiley, 1949.